



Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials

Citation

Murray, Eleanor J, and Miguel A Hernán. 2016. "Adherence Adjustment in the Coronary Drug Project: A Call for Better Per-Protocol Effect Estimates in Randomized Trials." *Clinical Trials: Journal of the Society for Clinical Trials* 13 (4) (March 7): 372–378. doi:10.1177/1740774516634335.

Published Version

10.1177/1740774516634335

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:34638499>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Dec 17, 2015

**Adherence adjustment in the Coronary Drug Project:
A call for better per-protocol effect estimates in randomized trials**

Eleanor J Murray¹, Miguel A Hernán^{1,2}

1. Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
2. Harvard-MIT Division of Health Sciences and Technology, Boston, MA

Funding: This work was partly funded by NIH grants R01 AI102634 and R01 HL080644.

Running head: Call for better per-protocol estimates

Word Count: 3575

Corresponding Author: Eleanor Murray, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115. Email: emurray@mail.harvard.edu

Trial registration number and trial register: ClinicalTrials.gov Identifier – NCT00000482

Abstract (word count: 191/450)

Background: In many randomized controlled trials, patients and doctors are more interested in the per-protocol effect than in the intention-to-treat effect. However, valid estimation of per-protocol effect generally requires adjustment for prognostic factors associated with adherence. These adherence adjustments have been strongly questioned in the clinical trials community, especially after 1980 when the Coronary Drug Project (CDP) team found that adherers to placebo had lower 5-year mortality than non-adherers to placebo.

Methods: We replicated the original CDP findings from 1980 and re-analyzed the CDP data using technical and conceptual developments that have become established since 1980. Specifically, we used logistic models for binary outcomes, decoupled the definition of adherence from loss to follow-up, adjusted for pre-randomization covariates via standardization and for post-randomization covariates via inverse probability weighting.

Results: The original CDP analysis reported a difference in 5-year mortality between adherers and non-adherers in the placebo arm of 9.4 percentage points. By using modern approaches, we found that this difference was reduced to 2.5 (95% CI -2.1, 7.0).

Conclusions: Valid estimation of per-protocol effects may be possible in randomized clinical trials when analysts use appropriate methods to adjust for post-randomization variables.

Keywords

Per-protocol effect, intention-to-treat effect, inverse probability weighting, Coronary Drug Project, adherence

Clinical decision making relies heavily on the findings of randomized clinical trials. Because substantial societal resources are allocated to the conduct of these studies, it is surprising that randomized trial findings are often summarized by a single measure that provides incomplete clinical information: the intention-to-treat effect. In an intention-to-treat analysis individuals are assigned to the treatment arm they were randomized to, regardless of whether they actually took the treatment as intended in the study's protocol. In placebo-controlled trials, patients assigned to treatment may stop taking it and those assigned to placebo may start taking treatment. As a result, the intention-to-treat effect will be closer to the null than the effect that would have been found if all patients had taken their assigned study treatment exactly as specified by the study protocol, i.e., the per-protocol effect, which is the effect under full adherence to the protocol.

The intention-to-treat effect has well known advantages.¹ However, patients and their doctors may be more interested in the per-protocol effect than in the intention-to-treat effect when making treatment decisions.² In trials that try to detect harm, a closer-to-the-null-by-design intention-to-treat effect is misleading and generally discouraged. In the increasingly frequent pragmatic trials, the benefits of randomization may be overwhelmed by serious deviations from protocol, including lack of adherence to the assigned interventions and loss to follow-up. Therefore, in many randomized trials, intention-to-treat analyses need to be complemented with properly conducted per-protocol analyses that adjust for incomplete adherence.³

However, there is no universally accepted method to estimate per-protocol effects. The naïve approach commonly referred to as “per-protocol analysis” (that is, comparing only individuals who adhere to their assigned treatment while they adhere to it) is hardly a contender for a valid estimation of the per-

protocol effect because this approach would only be valid if non-adherence occurred essentially at random.² Adjustment for pre-randomization variables will generally help reduce bias in naïve per-protocol analyses, but such adjustment is not common in practice. In fact, most clinical trial protocols devote little, if any, attention to the estimation of the per-protocol effect.

Some of this lack of interest can be traced back to an influential article published in 1980 the *New England Journal of Medicine*.⁴ This article had a chilling effect on subsequent attempts to conduct per-protocol analyses: investigators from the Coronary Drug Project (CDP) demonstrated that, among individuals assigned to placebo, the 5-year mortality risk was higher among those who did not adhere than among those who did adhere to the placebo pills. This finding is taught in courses around the world and often quoted as a reminder of the dangers of analyses that deviate from the intention-to-treat principle.

Because placebo cannot affect mortality, the difference between compliers and noncompliers in the CDP clearly indicates that adherence is a marker of good prognosis and that a comparison of adherers and non-adherers is inherently biased. That statistical adjustment, using the best methods available at that time, could not eliminate the bias was interpreted as an indication that analyses to estimate the per-protocol effect were to be distrusted, if not outright discouraged.

In the thirty-five years since the publication of the CDP paper, new statistical methods have been developed to adjust for protocol deviations.^{5,6} These methods can validly incorporate pre- and post-randomization variables; therefore they will generally result in less biased per-protocol effect estimates than those obtained from naïve per-protocol analyses. Randomized trials that, like the CDP, compare sustained interventions over long periods are the best candidates to benefit from these new methods, which estimate more robust per-protocol effect estimates, to complement the usual intention-to-treat analyses.

Here we (i) revisit the CDP findings in light of technical and conceptual developments that have become established since 1980, and (ii) review the options available for the estimation of per-protocol effects today. We start by summarizing the CDP and its findings.

A brief summary of the CDP

Design and conduct of the CDP

The CDP was a 6-arm double-blind placebo-controlled randomized clinical trial designed to identify treatments that would reduce premature mortality in men with a history of myocardial infarction (MI).⁷

⁸ The study was funded by the National Heart, Lung, and Blood Institute of the U.S. National Institutes of Health. Enrollment occurred between March 1966 and October 1969, and follow-up continued until late 1974 to early 1975. The trial compared five active treatments to placebo: clofibrate (ethyl chlorophenoxyisobutyrate, 1.8mg/day), low-dose equine estrogen (2.5mg/day), high-dose equine estrogen (5.0mg/day), dextrothyroxine (6mg/day), and nicotinic acid (3mg/day). Increased risk of adverse events was observed in the high- and low-dose estrogen arms and the dextrothyroxine arm; these arms were discontinued in 1970, 1973, and 1971, respectively.⁹

Men were eligible to participate if they were between 30-64 years of age at the time of randomization, had an ECG with documented MI at least 3 months before entry, a New York Heart Association functional class I (no limitation of physical activity) or II (slight limitation), no previous surgery for coronary artery disease, no other life-limiting diseases or conditions that could affect long-term follow-up, no contraindication for the study drugs, willingness to participate and to adhere to study drugs as assessed by a two-month control period, and no anti-coagulant therapy, lipid influencing drugs, or insulin use at baseline. Men were excluded if evidence suggested MI due to coronary artery embolism, aortic dissection, or prolonged arrhythmia.

Randomization was within each of 53 study centers and by risk group, with low risk defined as having had only one previous MI and no associated complications, and high risk as having had more than one MI or one MI with complications. Figure 1 displays an overview of the study timeline. Three initial monthly visits were conducted to determine eligibility and to estimate adherence to placebo; randomization occurred at the third visit, and dosage was adjusted at two additional monthly visits. Patients were prescribed three pills daily at randomization and dosage was increased by up to 3 pills per visit, on the basis of drug tolerance, to a final dose of nine daily pills by the fifth visit. Follow-up visits were planned to occur every four months until study completion, dropout, or death.⁷ Up to three close-out visits were planned after study completion to assess any impact of drug withdrawal. Three of the study arms (both estrogen arms and the dextrothyroxine arm) were discontinued during follow-up because of adverse events.¹⁰⁻¹²

Clofibrate versus placebo: Published results in 1980

1103 patients were randomized to the clofibrate arm and 2789 to the placebo arm.⁹ Minimum follow-up duration for surviving participants was 54 months; 96% were followed for at least 5 years and 63% for at least 6 years. As of August 31, 1974, 7.4% of survivors in the clofibrate arm and 8.0% of those in the placebo arm had stopped attending study visits and were classified as lost to follow-up. At 56-60 months, 5.5% of the clofibrate arm and 4.2% of the placebo arm were receiving less than 20% of the prescribed dose (p-value= 0.28), and 11.5% of the clofibrate arm and 9.4% of the placebo arm were receiving less than 80% of the prescribed dose (p-value =0.20). Mean adherence to the maximum study dose was 77.1% in the clofibrate arm and 77.8% in the placebo arm; median adherence was 86.1% for clofibrate and 87.1% for placebo.⁹ The percentage of patients with adherence of less than 80% by the end of follow-up was 5.5% for the clofibrate arm and 9.4% in the placebo arm.⁹

The 5-year all-cause mortality was 20.0% for the clofibrate arm and 20.9% in the placebo arm (p-value = 0.55). The five-year incidence for the secondary combined outcome of death due to coronary heart disease or occurrence of definite non-fatal MI was 23.8% for clofibrate and 26.2% for placebo (p-value = 0.13). Adjustment for baseline characteristics did not materially alter the study results.⁹ The CDP Research Group concluded that there was no evidence that clofibrate reduced 5-year mortality.⁹

How the Coronary Drug Project convinced us that adjustment for non-adherence is futile

In 1980 the CDP investigators also compared the 5-year mortality risk between adherers and non-adherers within each arm. To calculate adherence, the clinic staff computed the percentage of capsules prescribed that were actually taken by the patient during the previous four months. This assessment was made by questioning the patient and/or counting the capsules returned at each visit. Adherence was categorized into one of five levels: 80-100%, 60-79%, 40-59%, 20-39%, and less than 20% adherence to prescribed dosage. Adherence was standardized by prescribed dosage, and the cumulative adherence was computed from this value until death or the end of year 5. Finally, cumulative adherence was dichotomized at less than 80% versus greater than or equal to 80% over the follow-up.⁴

In the placebo arm, 882 participants had a cumulative adherence of less than 80% (noncompliers) and 1813 participants had a cumulative adherence of greater than or equal to 80% (compliers) until death or end of follow-up. Among noncompliers, 28% died within the first 5 years of follow-up, compared with only 15% of compliers; a difference of over 13%. After adjustment for 40 baseline characteristics, the difference in 5-year mortality risk was still about 10%. The authors concluded that compliers and non-compliers must be different in ways not accounted for by the available data because a greater adherence to placebo is not expected to have a causal effect on the mortality risk. This inability to adjust for adherence in the placebo arm strongly suggests that any attempt at an adherence-adjusted analysis

of the entire trial data would be futile. By extension, if adherence adjustment cannot be done in the CDP, a large trial with extensive and detailed data collection at baseline and over time, why should we expect to do so in any other randomized trial?

An early 21st century update of the CDP placebo comparisons

The approaches to data analysis that were standard in 1980 would not necessarily be the primary choice today. A modern comparison of compliers and noncompliers in the placebo arm of the CDP would differ from the 1980 analysis.

We conducted a re-analysis of the placebo arm of the CDP to estimate the difference in 5-year mortality risk under adherence levels of $<80\%$ versus $\geq 80\%$. Because placebo should have no effect on the 5-year mortality risk, a non-null estimate would suggest that compliers and non-compliers are irreconcilably different. On the other hand, a null estimate would increase confidence in our ability to sufficiently adjust for the differences between compliers and noncompliers when estimating the per-protocol effect.

Our re-analysis differs from the original 1980 analysis in three aspects: definition of adherence during missed visits, adjustment for baseline predictors of adherence, and adjustment for post-randomization predictors of adherence. We proposed four updates to the 1980 analysis.

1. Definition of adherence

When a participant missed a study visit, his adherence to the placebo pills was unknown and therefore needed to be imputed. The 1980 analysis set adherence to 0% at missed visits, even though individuals may have partly adhered to placebo during the period preceding the missed visit. Because individuals

were more likely to miss visits when they were about to die (mortality hazard ratio for those who did versus did not miss the previous visit was 1.8, 95% CI: 1.3 to 2.6), setting adherence to 0% may artificially increase the proportion of noncompliers among patients at high risk of death.

We updated the analysis by decoupling the classification of an individual as a noncomplier from his risk of death in the near future. To do so, we used past adherence history as an indicator of expected adherence by setting adherence at missed visits equal to adherence at the most recently attended visit, which may better reflect an individual's expected adherence. We censor individuals who miss more than three consecutive adherence measurements. See the Appendix for details.

2. Adjustment for pre-randomization risk factors

To adjust for baseline factors, the 1980 analysis used linear regression to model the 5-year mortality risk as a function of the baseline covariates, which were then evaluated at their average values in the study population. That was a typical choice. We updated the analysis to a logistic regression model followed by standardization by the baseline covariates.

3. Adjustment for post-randomization risk factors

Because adherence varies over time, an analysis adjusted for only baseline variables may miss important prognostic factors that impacted adherence during the follow-up but that were not present at the time of randomization. The use of this post-baseline information will generally improve adherence-adjusted analyses. Indeed not adjusting for post-randomization variables makes the extreme assumption that, conditional on the baseline variables, adherence after randomization is independent of all measured and unmeasured post-randomization variables. The 1980 analysis, however, could not adjust for any post-baseline, time-varying covariates because no methods had yet been developed for this purpose at the time. As a result, the CDP investigators could not take advantage of much of the data so carefully collected over the course of the study.

We updated the analysis by using inverse probability weighting⁶ to adjust for post-randomization risk factors. Because the weight assigned to an individual at time t is a function of the individual's risk factor and adherence history through t , the analysis naturally adjusts for post-randomization variables that are joint determinants of adherence and the outcome. This analysis makes the less extreme assumption that adherence post-randomization is independent of the unmeasured but not the measured post-randomization variables. If all such variables were correctly measured and modeled,^{6,13} the inverse probability weighted analysis would allow us to validly estimate the 5-year mortality that would have been observed in this trial if everyone in the placebo arm had adhered at least 80% to placebo compared with if no one in the placebo arm had adhered at least 80%. See the Appendix for more details.

4. Sensitivity analyses

We performed a variety of sensitivity analyses for dealing with missing adherence, including setting adherence to the cumulative average adherence up to each missed visit, setting adherence to 0% or 80% for up to 2 missed visits and then censoring, and setting adherence to half of the most recent available adherence value, with and without censoring. We also varied the definition of adherence by including or ignoring information about prescribed dosage, and changing the cut-point for dichotomous adherence to 70% and 90%. Finally, we included weights for censoring due to dropout, and permuted the covariates to assess the potential for over-fitting. Full results of the sensitivity analyses are provided in the Appendix.

Revisiting the CDP placebo comparisons

The CDP data tapes in the NIH data repository were corrupted in the 1990s and the data were irretrievable. Fortunately, Dr. Paul Canner, principal statistician of the CDP, had preserved a version of the de-identified data sets, which we used for our analyses. In collaboration with Dr. Canner, we will

make the datasets user friendly. We will then submit the data sets and their documentation to the NIH data repository.

Replication of the 1980 CDP findings

As a check of the integrity of the data, we first replicated the main tables and results of the 1980 CDP article.⁴ To do so, we mimicked the methods described in the article as closely as possible (see Appendix for a detailed description).

Of 2630 individuals with non-missing values of the baseline covariates, 68.8% had a cumulative average adherence to placebo of at least 80%. We found a difference in 5-year mortality risk of 14.3 percentage points in the unadjusted analysis and of 10.9 after adjustment for baseline covariates. These estimates are within 1.5 percentage points of the estimates reported in the original 1980 analysis (unadjusted difference = 13.1%; adjusted difference = 9.4%). These small discrepancies are explained by the need to recalculate percent adherence and cumulative average adherence because of corruption in the original variables, and to correct some errors in coding for adherence at baseline and time of death. A similar analysis conducted in the clofibrate arm also showed a slight increase in 5-year mortality in the non-adherers compared to the original 1980 estimates (Table 1).

A 2015 re-analysis of the CDP

As described above, our updated analysis differs from the original CDP analysis in three aspects: (1) handling of adherence at missed visits, (2) type of regression model for adjustment using baseline predictors, and (3) adjustment for post-randomization predictors of adherence using inverse probability weighting. Changing the handling of adherence at missed visits reduced our sample size by 217 individuals. These individuals missed more than three consecutive visits and were censored.

Adherence of less than 80% was associated with low prior adherence; acute coronary insufficiency, intermittent cerebral ischemic attack, congestive heart failure, and angina; recent use of diuretics and anti-arrhythmic agents; and smoking (Appendix Table 2). The adherence weights, truncated at their 99th percentile to prevent undue influence of outliers, had a mean of 1.01 (SD = 0.51) and a range of 0.02 to 3.94.

Table 2 shows the estimated difference in 5-year mortality risk between noncompliers and compliers in the placebo arm of the CDP, both in our replication of the 1980 analysis and in our updated analysis. The unadjusted difference in 5-year mortality risk (95% CI) was 11.0 percentage points (6.5, 15.6). The difference in 5-year mortality risk was 7.0 percentage points (2.7, 11.2) after adjustment for baseline covariates. Our handling of adherence and the use of logistic regression resulted in a 2.4 percentage point reduction compared with the CDP analysis that set missing adherence to 0 and used linear regression. The difference in 5-year mortality risk was 2.5 percentage points (-2.1, 7.0) after further adjustment for post-randomization covariates.

That is, after adjustment, there was little survival difference between compliers and noncompliers in the placebo arm. This finding was robust to many possible variations of the analysis, which we explored in multiple sensitivity analyses (see Appendix).

Where does this leave us when estimating per-protocol effects?

The above shows that an analysis of the CDP that tries to adjust for adherence is not necessarily doomed. Our finding weakens the arguments against adherence-adjusted analyses of randomized trials for estimating the per-protocol effect. The question remains of how to better estimate such effect.

The two broad approaches to the estimation of per-protocol effects---instrumental variable estimation and direct adjustment for measured covariates--- each have shortcomings. Adjustment for noncompliance via instrumental variable estimation¹⁴ is desirable because it removes the need to collect data on prognostic factors that predict compliance. For example, Holme et al.¹⁵ recently used instrumental variable estimation to estimate the per-protocol effect of once-only sigmoidoscopy on colon cancer and overall mortality rate in a large randomized trial conducted in Norway. However, instrumental variable estimation requires extreme assumptions in trials with interventions that, unlike a baseline sigmoidoscopy, need to be sustained during the follow-up. The use of instrumental variable estimations is even more questionable if the interventions are unmasked and potentially available to individuals in all arms (not a concern for the CDP).

Direct adjustment for measured covariates is another approach for per-protocol analyses of randomized trials of sustained interventions. Typically, these analyses require adjustment for pre- and post-randomization prognostic factors that predict adherence. Appropriate adjustment for post-randomization factors requires methods specifically designed for that purpose. These methods, collectively referred as g-methods and developed by Robins and collaborators,^{13, 16} include inverse probability weighting, which we used in our updated analysis of the CDP.

The key limitation of direct adjustment is the possibility of residual bias if the set of adjustment variables fails to include important joint predictors of adherence to the protocol and the outcome of interest. (Note that when the analysis fails to adjust for important joint predictors of both loss to follow-up and of the outcome of interest, then intention-to-treat effect estimates may also be biased because of informative censoring).¹⁷ In fact, there are examples, other than the CDP, in which adherers and non-adherers are so different with respect to their risk of developing the outcome that the available data are insufficient for direct statistical adjustment. For example, Holme et al. demonstrated that the mortality rate was greater in participants who refused to undergo a sigmoidoscopy compared with those who did

undergo sigmoidoscopy, even after adjustment for multiple factors.¹⁵ However, the adjustment variables did not include cigarette smoking and other important prognostic factors, which, had they been available for inclusion might have explained the some of the observed difference in mortality rate.

Many randomized trials collect only limited data after randomization, which prevents valid estimation of per-protocol effects when instrumental variable estimation is not possible. Even in randomized trials that collect abundant baseline and post-randomization data, adjusted per-protocol effect estimates are not routinely presented along with intention-to-treat effect estimates. Appropriately adjusted per-protocol effect estimates can provide better estimates of both the efficacy of a treatment and the harms associated with the treatment in randomized trials. The omission of appropriately adjusted per-protocol effect estimates is especially problematic in randomized trials that assess potential harm and in those in which little effort is made to encourage adherence to the protocol, such as pragmatic and large simple trials.¹⁸ The lack of established methodologic standards for adjusted per-protocol estimates contributes to the problem because it creates uncertainty among investigators and sponsors who need to justify their analytic choices to journal editors and regulators.¹⁹

Conclusions

Widely held beliefs about the impossibility of conducting adherence-adjusted analyses may be partly based on weak empirical evidence and out-of-date methodology. Per-protocol effects can be validly estimated in clinical trials when sufficient data are available and appropriate adjustment methods are used. Supplementing intention-to-treat analyses with per-protocol effect estimates can allow richer, more nuanced inference from clinical trials. Clinical research would benefit from a concerted effort to establish clear guidelines for the definition and estimation of per-protocol effects.

Acknowledgments

The authors are indebted to Paul Canner, Principal Statistician of the Coronary Drug Project, and Roger Logan for his assistance. We thank Dean Follmann for his encouragement and assistance in the earlier stages of this project.

Conflict of interest statement: The Authors declare that there is no conflict of interest.

References

1. Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res* 2011; 2: 109-112.
2. Hernán MA and Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012; 9: 48-55.
3. Hernán MA, Hernandez-Diaz S and Robins JM. Randomized trials analyzed as observational studies. *Ann Intern Med* 2013; 159: 560-562.
4. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980; 303: 1038-1041.
5. Mark SD and Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Control Clin Trials* 1993; 14: 79-97.
6. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; 56: 779-788.
7. Coronary Drug Project Research Group. The coronary drug project. Design, methods, and baseline results. *Circulation* 1973; 47: 11-50.
8. Coronary Drug Project Research Group. Factors influencing long-term prognosis after recovery from myocardial infarction--three-year findings of the coronary drug project. *J Chronic Dis* 1974; 27: 267-285.
9. Coronary Drug Project Research Group. Clofibrate and niacin in coronary heart disease. *JAMA* 1975; 231: 360-381.
10. Coronary Drug Project Research Group. The Coronary Drug Project. Findings leading to discontinuation of the 2.5-mg day estrogen group. *JAMA* 1973; 226: 652-657.
11. Coronary Drug Project Research Group. The Coronary Drug Project. Findings leading to further modifications of its protocol with respect to dextrothyroxine. *JAMA* 1972; 220: 996-1008.
12. The Coronary Drug Project. Initial findings leading to modifications of its research protocol. *JAMA* 1970; 214: 1303-1313.
13. Robins JM and Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G and Molenberghs G (eds). *Longitudinal data analysis*. New York: Chapman and Hall/CRC Press, 2008, pp. 553-599.
14. Hernan MA and Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17: 360-372.
15. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA* 2014; 312: 606-615.
16. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986; 7: 1393-1512. *Errata in Computers and Mathematics with Applications* 1987; 14: 917-921.
17. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; 367: 1355-1360.
18. Eapen ZJ, Lauer MS and Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. *JAMA* 2014; 311: 1397-1398.
19. US Food and Drug Administration. Guidance for industry: determining the extent of safety data collection needed in late stage premarket and postapproval clinical investigations (draft guidance), <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291158.pdf> (2012, accessed 16 November 2015)

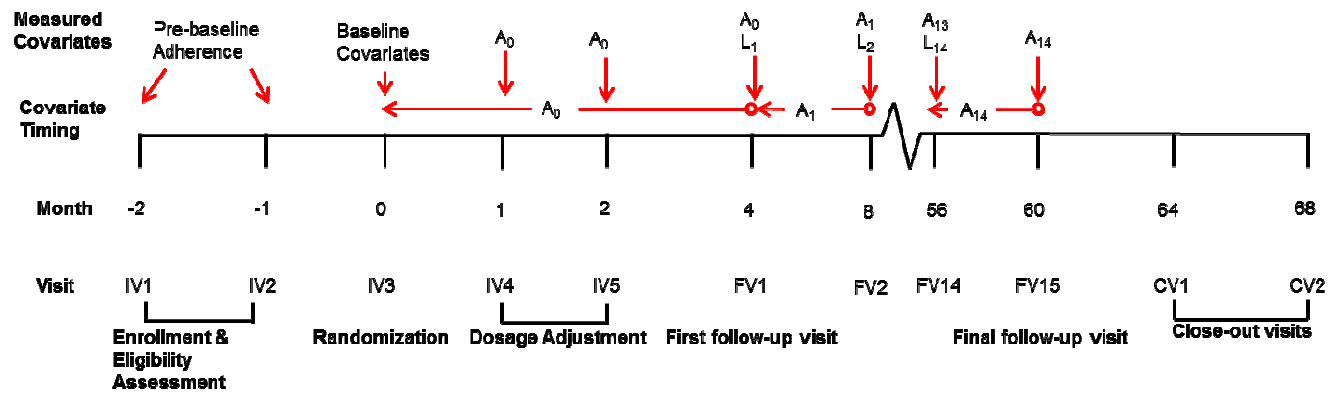


Figure 1: Study timeline, Coronary Drug Project. Timing of study visits and variable measurement. L_t is the vector of post-randomization covariates measured at visit t and A_t is an indicator for adherence level for the period from t to just before $t+1$ measured at visit $t+1$. A_0 is measured at follow-up visit 1 (FV1) or, for individuals who died or dropped out before FV1, at IV4 and IV5.

Table 1: Replication of the original results reported in reference 4, Coronary Drug Project

	5-year mortality, % (95% CI)			
	Unadjusted		Adjusted (linear regression with standardization to baseline covariates)	
	Cumulative Adherence < 80%	Cumulative Adherence ≥80%	Cumulative Adherence < 80%	Cumulative Adherence ≥80%
Original 1980 analysis, Placebo arm ⁴	28.2 (25.3, 31.1)	15.1 (13.5, 16.7)	25.8 (22.9, 28.7)	16.4 (14.8, 18.0)
Replication of original 1980 analysis, Placebo arm	29.7 (26.6, 33.0)	15.5 (13.8, 17.2)	27.4 (24.5, 30.3)	16.5 (14.8, 18.2)
Original 1980 analysis, clofibrate arm ⁴	24.6 (20.1, 29.1)	15.0 (12.5, 17.5)	22.5 (18.0, 27.0)	15.7 (13.2, 18.2)
Replication of original 1980 analysis, clofibrate arm	26.1 (21.5, 31.2)	15.3 (12.7, 18.2)	23.8 (19.7, 27.8)	16.4 (13.7, 19.2)

Table 2. Comparison of original and updated estimates for the placebo arm, Coronary Drug Project

	5-year mortality risk difference, % (95% CI)			
	Unadjusted	Adjusted for baseline variables* via linear regression	Adjusted for baseline variables* via logistic regression+	Further adjusted for post- randomization variables**
Replication of the original 1980 analysis (N = 2630)	14.3 (10.8, 17.8)	10.9 (7.5, 14.4)	10.6 (7.3, 14.0)	N/A
Updated 2015 analysis (N = 2401)	11.0 (6.5, 15.6)	7.4 (3.0, 11.8)	7.0 (2.7, 11.2)	2.5 (-2.1, 7.0)

*Adjusted for 39 baseline variables: age; race; risk group; number of prior myocardial infarctions; relative body weight; medical history; prescriptions of non-study medications; lab findings; blood pressure; cardiomegaly; ECG findings; cigarette smoking; and physical activity level (See Appendix Table 1 for details).

**Adjusted for the 39 baseline variables (age, race, risk group, prior MIs, and relative body weight were baseline only) and 34 post-randomization variables: medical history; prescriptions of non-study medications; lab findings; blood pressure; cardiomegaly; ECG findings; cigarette smoking; and physical activity level (See Appendix Table 1 for details). Comparing individuals with <80% versus ≥80% at each visit.

+The original 1980 analysis did not include logistic regression; we include it here for comparison purposes

Dec 17, 2015

**Adherence adjustment in the Coronary Drug Project:
A call for better per-protocol effect estimates in randomized trials**

Eleanor J. Murray¹, Miguel A. Hernán^{1,2}

1. Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
2. Harvard-MIT Division of Health Sciences and Technology, Boston, MA

Funding: This work was partly funded by NIH grants R01 AI102634 and R01 HL080644.

Corresponding Author: Eleanor Murray, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115. Email: emurray@mail.harvard.edu

Trial registration number and trial register: ClinicalTrials.gov Identifier – NCT00000482

Appendix

1. Compliers vs noncompliers in the placebo arm: Replication of the 1980 CDP results

Individuals were followed from randomization until the end of follow-up, which was defined as the earliest of 60 months, end of study (August 31, 1974), or death. Individuals who stopped attending study visits were coded as dropouts but not censored; vital status on August 31, 1975 was known for all but 1 dropout in the placebo arm and everyone in the clofibrate arm [1]. Of the 1103 individuals in the clofibrate arm, 311 died and 228 dropped out. Of the 2787 individuals in the placebo arm, 780 died and 546 dropped out. Excluding 128 patients who started follow-up after August 31, 1969 and therefore were not followed for the full 5 years did not materially change the estimates below.

Adherence

CDP participants were assessed for adherence to placebo at dosage adjustment visits at month 1 and 2 after randomization. Baseline adherence measure is taken as the average of percent adherence to placebo at these visits (IV4 and IV5). Adherence at each follow-up visit (IADH1-26) was classified as at least 80%, 60-79%, 40-59%, 20-39%, or less than 20% of the study dose over the previous 4-month period. Visits with missing adherence or prescription information because of missed visits, incomplete forms, or drop-out, and with no capsules prescribed, were assigned adherence equal to zero. Adherence was dose-adjusted by the percent of study dose prescribed during the previous 4-month period. Cumulative average adherence was then computed from percent adherence to prescription over the entire follow-up. Further details on the adherence measures can be found in the CDP documentation.

Covariates

We used the 40 covariates included in the 1980 paper [2] (Table A1), except “time since most recent MI”, which was missing for nearly half the sample (1123/2787 placebo arm participants, 421/1103 clofibrate arm participants) and had values less than 999 days for all others because of an error with missing indicator levels at some point in the past.

The baseline covariates used in the 1980 analysis were selected from the full set of approximately 135 baseline characteristics recorded at study entry by the CDP study team on the basis of their status as known or suspected risk factors for first MI [2, 3]. More information about these covariates and the selection process, as well as their relationships with recurrent MI among the CDP participants, was reported by the study team in 1974 [3].

Of the 39 covariates, 5 were recorded at baseline only and 34 were measured at baseline and during follow-up. Of the 34, 20 were recorded at every follow-up visit, while 14 were only recorded at annual-type visits which occurred at FV3, 6, 9, 12, and 15. Many additional variables were available in the CDP data both at baseline and during follow-up. We limited our analyses to the variables used in the original 1980 paper for the sake of comparability between our and previous results.

The 1980 paper did not specify whether the baseline covariates were dichotomized as in the main clofibrate analysis [1] or unchanged from the original coding as in interim analyses, in which the 40 covariates were selected [3]. However, Dr. Canner recalls that the 1980 paper dichotomized all covariates so we did the same.

Statistical Analysis

To adjust for baseline variables, we fit the linear regression model $E[Y|cumA] = \beta_0 + \beta_1 cumA + \beta_2 V$ where Y is an indicator for death by the end of follow-up (1: yes, 0: no), $cumA$ is an indicator for at least 80% cumulative average adherence to placebo by the end of follow-up (1: yes, 0: no), and V is a vector including the 39 dichotomized baseline covariates to the model. The predicted values of this model for $cumA=1$ and $cumA=0$, standardized for the baseline covariates in V , were our estimates of the 5-year mortality risks. We used standardization rather than conditioning on a particular combination of values of the covariates because the estimates were sensitive to the choice of values and it is not clear what values the 1980 paper used. The 95% confidence intervals were obtained via nonparametric bootstrap with 500 samples.

2. Compliers vs noncompliers in the placebo arm: Re-Analysis of CDP

Handling of adherence at missed visits

Instead of setting adherence level to 0 at some visits like the 1980 analysis did, we carried forward adherence from the most recent attended visit for up to two consecutive missed visits. Individuals who missed three or more consecutive visits were censored at the third consecutive missed visit (Figure A1). Under this scheme, 217 individuals were censored from all analyses because of missing adherence data, and 3 additional individuals were censored from the post-randomization adjusted analysis because of missing covariate information (Table A3).

Type of regression model for adjustment using baseline predictors

Instead of using linear regression for the outcome model adjusted for baseline variables, we used logistic regression because the outcome (death within 5 years) is binary. The model was $\text{logit}(\Pr[Y=1 | cumA, V, C_{15}=0]) = \beta_0 + \beta_1 cumA + \beta_2 V$, where Y , $cumA$, and V are defined as above, and C_{15} is an indicator of censoring (1: yes, 0: no) by the end of follow-up at visit 15. The predicted values of this model for $cumA=1$ and $cumA=0$, standardized the estimates over the baseline covariates V , were our estimates of the 5-year mortality risks. Nonstandardized estimates when covariates were set to their median value were similar (risk difference: 4.1, 95% CI: 1.1, 7.2).

Adjustment for post-randomization predictors of adherence

To adjust for post-randomization predictors of adherence we used inverse probability (IP) weighting [4] for adherence at time t and applied the weights to the same outcome model used for baseline covariate adjustment: $\text{logit}(\Pr[Y=1 | cumA, V, C_{15}=0]) = \beta_0 + \beta_1 cumA + \beta_2 V$.

The weights were estimated as follows. Let t be an index for visit number, where $t=0$ is the baseline visit. Each uncensored subject contributing to the outcome model received a stabilized IP weight $SW = \prod_{t=0}^{14} \frac{f[M_t, A_t | A_0, V, \bar{A}_{t-1}, C_t = 0]}{f[M_t, A_t | A_0, V, \bar{A}_{t-1}, \bar{L}_t, C_t = 0]}$, where M_t is an indicator for measurement of adherence at visit t (1 if measured, 0 otherwise), A_t is the adherence level between visits t and $t+1$ (0: $\geq 80\%$, 1: $< 80\%$), \bar{A}_{t-1} is adherence history through t , V is a vector of baseline covariates, \bar{L}_t is covariate history through t , and $C_t=0$ is an indicator for remaining in the study at visit t .

To estimate the denominator of the weights, we fit models in two stages. First, we fit the logistic model $\text{logit}(\Pr[M_t=1 | A_0, V, \bar{A}_{t-1}, \bar{L}_t, C_t = 0]) = \beta_{0t} + \beta_1 A_0 + \beta_2 V + \beta_3 \bar{A}_{t-1} + \beta_4 \bar{L}_t$ to all person-visits. Second, we fit the logistic model $\text{logit}(\Pr[A_t=0 | A_0, V, \bar{A}_{t-1}, \bar{L}_t, C_t = 0, M_t=1]) = \beta_{0t} + \beta_1 A_0 + \beta_2 V + \beta_3 \bar{A}_{t-1} + \beta_4 \bar{L}_t$ to all person-visits with measured adherence. Visits when adherence was not directly measured ($M_t=0$), but carried forward from the previous visit, the factor in the denominator of the adherence weight was 1 for that visit.

Similar models that did not include the time-varying covariates were fit to estimate the numerators of the weights. The final weight for each individual was the product of the measurement and adherence weights for that individual at each time point until censoring. As in previous studies[5, 6] we truncated the estimated IP weights at the 99th percentile to avoid undue influence of outliers. In our application, truncation of the weights did not materially change the estimates. For all analyses, 95% confidence intervals were obtained via nonparametric bootstrap with 500 samples.

Adjustment for censoring due to dropout

In a sensitivity analysis to adjust for potential bias from censoring due to dropout, we estimated stabilized IP censoring weights SW^C

$= \prod_{t=0}^{14} \Pr[C_t = 0 | A_0, V, \bar{A}_{t-1}, C_{t-1} = 0] / \Pr[C_t = 0 | A_0, V, \bar{A}_{t-1}, \bar{L}_t, C_{t-1} = 0]$ by fitting the logistic regression model $\text{logit}(\Pr[C_t=0 | A_0, V, \bar{A}_{t-1}, \bar{L}_t, M_t = 1]) = \beta_{0t} + \beta_1 A_0 + \beta_2 V + \beta_3 \bar{A}_{t-1} + \beta_4 \bar{L}_{t-1}$ to visits when adherence had been missing for the previous two visits; at all other times the conditional probability of remaining in the study is 1 for all previously uncensored individuals.

Similar models that did not include the time-varying covariates were fit to estimate the numerators of the stabilized censoring weights. The final weights for the censoring-adjusted models were the product of the censoring weights and the previously described adherence weights over all time points. To avoid undue influence of outliers, we again truncated the final weights at the 99th percentile.

Adding the inverse probability of censoring weights did not materially change the estimates; the estimated post-randomization-adjusted difference in 5-year mortality was 2.5% (-3.3, 8.8) after adjustment for censoring. The estimated censoring weights had a mean of 0.99 (SD = 0.06) and a range of 0.16 to 1.00. The combined adherence and censoring weights had a mean of 1.00 (SD = 0.50) and a range of 0.02 to 3.94.

Sensitivity Analyses

In our main analysis, we carried adherence forward from the previous visit for up to two consecutive missed visits and then censored individuals on their third consecutive missed visit. We also assessed other methods of dealing with missing adherence, which we present here and in Table A4. For all sensitivity analyses, 95% confidence intervals were obtained via bootstrap with 500 samples.

We ran 6 sensitivity analyses for dealing with missing adherence, all but one of which resulted in a null effect of adherence in the post-randomization adjusted analyses. When adherence was carried forward for two missed visits and set to 0 only from the third missed visit, instead of censoring the individual, the estimated difference in mortality at 5-years was 0.2% (-3.7, 4.1) after adjustment for post-randomization variables. When adherence was carried forward for two missed visits and set to 80% from the third missed visit, the estimated post-randomization-adjusted difference was 0.9% (-3.2, 4.9). When

adherence was set to the cumulative average adherence reported prior to any missed visit, with no censoring, the estimated post-randomization-adjusted difference at 5 years was 1.1% (-3.4, 5.7).

When missing adherence was set to half that at the last available measurement, the estimated post-randomization adjusted difference was 1.4% (-2.6, 5.4) with no censoring, and 3.6% (-1.0, 8.1) with censoring at the third consecutive missed visit. When adherence was set to 0 for up to two missed visits and censored from the third consecutive missed visit (similar to 1980 analysis but with censoring after 3 missed visits) the estimated difference was 3.2% (-1.2, 7.5). These last two approaches are less successful at breaking the link between missing adherence and mortality and thus give estimates further from the null.

When adherence was set to 0 at all missed visits (the 1980 approach), the post-randomization-adjusted 5-year mortality difference was 7.6% (2.0, 13.1). We assessed the effect of definition of adherence in two ways: by ignoring prescribed dosage when calculating adherence, and by varying the cut-point for adherent versus non-adherent. Ignoring prescribed dosage produced estimates that were further from the null than the main analyses, particularly in the analyses adjusted for post-randomization variables and censoring which resulted in a 5-year mortality risk difference of 6.3% (-2.7, 15.2). We also varied the cut-point for binary cumulative adherence to 70% and to 90% for the unadjusted and baseline adjusted analyses. We were unable to adjust the cut-point for binary adherence at each visit because of the way this data were collected, so we were not able to run the post-randomization adjusted model at these levels. When cumulative adherence was dichotomized at 70%, the 5-year mortality difference was 8.2 percentage points (1.9, 14.4) in the baseline adjusted model. When cumulative adherence was dichotomized at 90% the 5-year mortality difference was -2.3 percentage points (-5.4, 0.8) in the baseline adjusted model. This latter finding may be explained because of the way the adherence data were coded by Canner with the midpoint of each category used for calculating cumulative adherence. Individuals with cumulative average adherence of 90% by the end of follow-up must have reported adherence in the 80-100% category at every visit while, and this was less common the longer an individual was under follow-up.

Some individuals were randomized after Aug 31, 1969 and therefore were administratively censored before completing the full 5 years of follow-up; excluding these individuals did not appreciably change the estimated difference in 5-year mortality risk (2.7%, -4.0, 9.3).

Finally, to check for over-fitting of the post-randomization adjusted analyses due to the large number of covariates in the model, we permuted the values of all baseline and time-varying covariates, while maintaining the link between adherence and 5-year mortality. We permuted the data 500 times, re-running the analyses on each, and obtained the mean, 2.5th and 97.5th percentiles of the distribution. In the permuted dataset, all covariates are independent of adherence and survival. As expected, analysis of this permuted dataset resulted in estimates similar in size to the unadjusted analysis. The mean baseline adjusted risk difference was 11.1 percentage points (2.5th, 97.5th percentiles: 10.5, 11.8), and the mean post-randomization-adjusted risk difference was 11.3 percentage points (2.5th and 97.5th percentiles: 10.1, 12.3). These results suggest little or no role of over-fitting of the post-randomization adjusted models in the findings.

SAS code

We have provided the SAS code used for the main analyses. SAS 9.4 was used for all analyses. The code appendix contains the following programs:

1. Program 1: Data management
This program reads in the CDP dataset and defines the variables for the original and updated analyses. It creates two datasets for the placebo arm. Changing the value of ITR to 3 produces the same datasets for the clofibrate arm.
 - Binary.sas7bdat for crude and baseline adjusted analyses
 - Censwt_ag.sas7bdat for post-randomization adjusted analyses
2. Program 2: Replication of 1980 analysis
This program estimates the crude and baseline adjustment using the original 1980 definition of adherence. The output gives the first 2 columns of row 1 of Table 2 when using the original adherence coding and the second column of row 2 when using the updated adherence coding. It also outputs the results for Table 1.
3. Program 3: Updated analyses with baseline variables
This program estimates the crude and baseline adjustment using updated definition of adherence and outputs columns 1 and 3 of row 2 of Table 2. Changing the adherence variable input allows estimation of the third column of row 1 in Table 2.
4. Program 4: Updated analysis with post-randomization covariates
This program requires the restricted cubic spline macro developed by F.E. Harrell[7, 8] in order to run. It estimates the association between adherence and death using adjustment for post-randomization covariates using inverse probability weighting and outputs final column for row 2 of Table 2.

References

1. Coronary Drug Project Research Group. Clofibrate and niacin in coronary heart disease. *JAMA*. 1975;231(4):360-81.
2. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med*. 1980;303(18):1038-41.
3. Coronary Drug Project Research Group. Factors influencing long-term prognosis after recovery from myocardial infarction--three-year findings of the coronary drug project. *J Chronic Dis*. 1974;27(6):267-85.
4. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56(3):779-88.
5. Cain LE, Logan R, Robins JM, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med*. 2011;154(8):509-15.
6. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008;168(6):656-64.
7. Harrell FE. %rcspline macro. Clinical Biostatistics, . Duke University Medical Center, 1988.
8. Devlin TF, Weeks BJ. Spline functions for logistic regression modeling. In: *Proc Eleventh Annual SAS Users Group International*, Atlanta, Georgia, February 9-12 1986: Cary NC: SAS Institute.

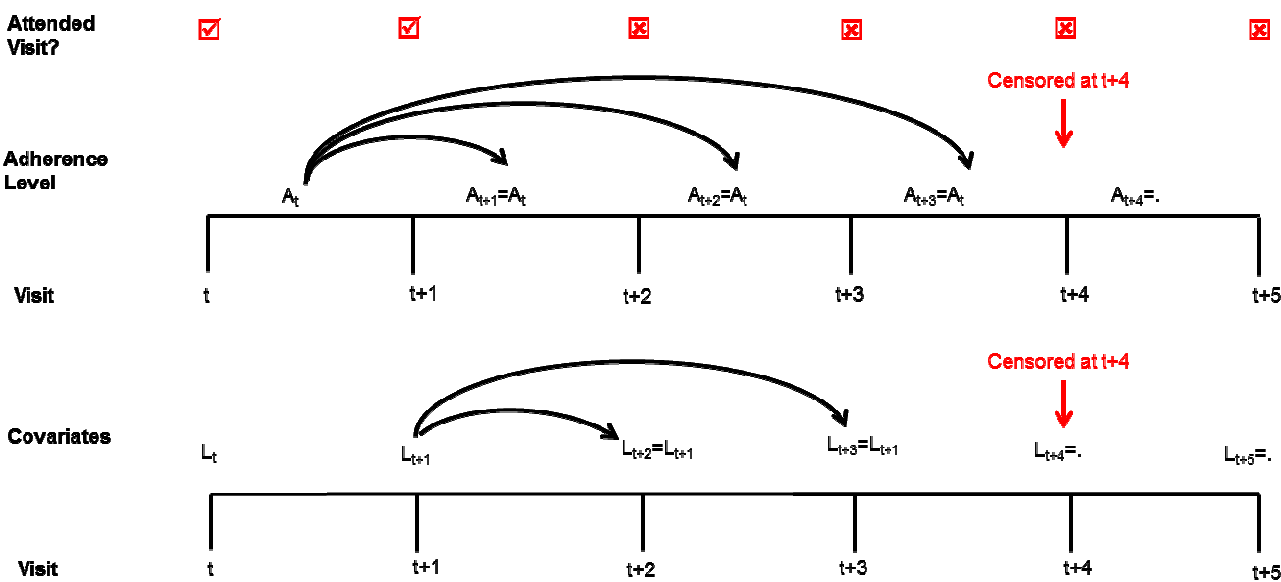


Figure A1: Censoring scheme for re-analyses of the Coronary Drug Project. L_t is the vector of covariates measured at visit t and A_t is an indicator for adherence level for the period from t to just before $t+1$. When an individual misses a study visit, the most recent covariate and adherence values are carried over from the previous inter-visit period. An individual is censored at the date of their third consecutive missed study visit. Covariates are therefore carried forward up to two consecutive missed visits, and adherence is carried forward up to three intervals between missed visits.

Table A1: Covariates used for adjustment, Coronary Drug Project

Covariate	Variable Code	Measured at:	Levels when dichotomized
<i>Demographics:</i>			
Age at entry (years)	IAGE	Baseline only	≥55 vs <55
Race	IRC	Baseline only	White vs not white
<i>Lifestyle characteristics:</i>			
Cigarette smoking	ICG1-6	Annual visits only	Any/none
Current habitual level of physical activity in leisure time	IJOB34-39	Annual visits only	“Vigorous or Moderate” vs “Light or Sedentary”
<i>Medical history:</i>			
Risk group	IRK	Baseline only	“Low risk (1 previous MI with no complications)” vs “High risk (>1 previous MI or MI with complications)”
Number of myocardial infarctions at baseline	NMI	Baseline only	≥2 vs <2
Time since most recent myocardial infarction	LMI	Baseline only	Dropped because of high missingness
Relative body weight	RBW	Baseline only	≥1.15 vs <1.15
History of congestive heart failure (yes = suspect or definite)	KLIN1; ICCL1-26	All visits	Ever at Baseline: yes/no Since last visit: yes/no
History of angina pectoris (yes = suspect or definite)	KLIN2; ICCL27-52	All visits	Ever at Baseline: yes/no Since last visit: yes/no
History of acute coronary insufficiency (yes = suspect or definite)	KLIN3; ICCL53-78	All visits	Ever at Baseline: yes/no Since last visit: yes/no
History of intermittent cerebral ischemic attack (yes = suspect or definite)	KLIN5; ICCL157-182	All visits	Ever at Baseline: yes/no Since last visit: yes/no
History of intermittent claudication (yes = suspect or definite)	KLIN7; ICCL235-260	All visits	Ever at Baseline: yes/no Since last visit: yes/no
Systolic blood pressure (mmHg)	WTPB32-47	All visits	≥130 vs <130
Diastolic blood pressure (mmHg)	WTPB61-76	All visits	≥85 vs <85
New York Heart Association functional class	NIHA1-11	Annual visits only	“Any limitation” vs “no limitation”
Cardiomegaly on chest X-ray (yes = probable or definite)	IXRAY1-22	Annual visits only	Yes/no
<i>Prescription of non-study medications:</i>			
Oral hypoglycemic agents	IMDCT29-56	All visits	Current at baseline: yes/no Since last visit yes/no
Digitalis	IMDCT57-84	All visits	Current at baseline: yes/no Since last visit yes/no
Antiarrhythmic agents	IMDCT85-112	All visits	Current at baseline: yes/no Since last visit yes/no
Diuretics	IMDCT113-140	All visits	Current at baseline: yes/no Since last visit yes/no
Antihypertensives other than diuretics	IMDCT141-168	All visits	Current at baseline: yes/no Since last visit yes/no
<i>Lab findings:</i>			
Serum total bilirubin (mg/dl)	ZLABB1; ZLAB4-18	All visits	≥0.5 vs <0.5
Serum total cholesterol (mg/dl)	ZLABB4; ZLAB91-105	All visits	≥250 vs <250
Serum triglyceride (meq/L)	ZLABB5;	All visits	≥5.0 vs <5.0

	ZLAB120-134		
Serum uric acid (mg/dl)	ZLABB6; ZLAB149-163	All visits	≥ 7.0 vs < 7.0
Serum alkaline phosphatase (K-A units)	ZLABB7; ZLAB178-192	All visits	≥ 7.5 vs < 7.5
Plasma urea nitrogen (mg/dl)	ZLABB8; ZLAB207-221	All visits	≥ 16 vs < 16
Plasma fasting glucose (mg/dl)	ZLABB9; ZLAB236-250	All visits	≥ 100 vs < 100
Plasma one-hour glucose, after 75g oral load (mg/dl)	ZLABB10; ZLAB265-279	All visits	≥ 180 vs < 180
White blood cell count (cells/mm ³)	WBC1-11	Annual visits only	≥ 7500 vs < 7500
Absolute neutrophil count (cells/mm ³)	WBC12-22	Annual visits only	≥ 4500 vs < 4500
Hematocrit (%)	WBC23-33	Annual visits only	≥ 46 vs < 46
<i>ECG findings:</i>			
Q/QS pattern on antero-lateral, postero-inferior, or antero-septal recording	IECG1-36	Annual visits only	Any/None
ST segment depression on antero-lateral, postero-inferior, or antero-septal recording	IECG37-72	Annual visits only	Any/None
T-wave findings on antero-lateral, postero-inferior, or antero-septal recording	IECG73-108	Annual visits only	Any/None
ST segment elevation on antero-lateral, postero-inferior, or antero-septal recording	IECG109-144	Annual visits only	Any/None
Ventricular conduction defect	IECG193-204	Annual visits only	Any/None
Premature ventricular beats (frequent ventricular ectopic beats) (Note: any does not include premature atrial or nodal beats)	IECG205-216	Annual visits only	Any/None
Heart rate (beats per minute)	IECG277-88	Annual visits only	≥ 70 vs < 70

Table A2: Predictors of low adherence (<80%), Coronary Drug Project

Covariate	OR (95% CI)
<i>Adherence history:</i>	
Low adherence	
pre-baseline	1.32 (1.11, 1.58)
at baseline	1.18 (1.01, 1.37)
at most recent visit	23.40 (21.12, 25.92)
<i>Demographics:</i>	
Age ≥ 55 years at baseline	1.05 (0.94, 1.16)
Race (not white)	0.96 (0.78, 1.17)
<i>Lifestyle characteristics:</i>	
Cigarette smoking	
at baseline	1.30 (1.11, 1.51)
at most recent visit	0.98 (0.84, 1.15)
Current habitual level of physical activity in leisure time (light or sedentary)	
at baseline	0.90 (0.81, 1.01)
at most recent visit	1.03 (0.92, 1.15)
<i>Medical history:</i>	
Risk group (>1 previous MI or MI with complications)	0.96 (0.83, 1.11)
Number of prior MIs (≥ 2)	0.96 (0.81, 1.14)
Relative body weight (≥ 1.15)	1.04 (0.94, 1.15)
Congestive heart failure	
at baseline	1.13 (0.96, 1.32)
at most recent visit	2.14 (1.73, 2.66)
Angina pectoris	
at baseline	0.98 (0.88, 1.11)
at most recent visit	1.22 (1.08, 1.36)
Acute coronary insufficiency	
at baseline	0.91 (0.799, 1.04)
at most recent visit	4.97 (3.96, 6.24)
Intermittent cerebral ischemic attack	
at baseline	1.18 (0.92, 1.52)
at most recent visit	2.31 (1.56, 3.42)
Intermittent claudication	
at baseline	1.06 (0.88, 1.27)
at most recent visit	1.03 (0.85, 1.26)
Systolic blood pressure (≥ 130 mmHg)	
at baseline	1.07 (0.95, 1.21)
at most recent visit	0.98 (0.87, 1.10)
Diastolic blood pressure (≥ 85 mmHg)	
at baseline	1.04 (0.92, 1.18)
at most recent visit	1.03 (0.91, 1.16)
New York Heart Association functional class (any limitation)	
at baseline	0.98 (0.87, 1.11)
at most recent visit	1.05 (0.93, 1.19)
Cardiomegaly on chest X-ray	
at baseline	0.96 (0.83, 1.12)
at most recent visit	1.04 (0.90, 1.20)
<i>Prescription of non-study medications:</i>	
Oral hypoglycemic agents	
at baseline	0.96 (0.69, 1.33)
at most recent visit	0.86 (0.64, 1.16)
Digitalis	
at baseline	0.73 (0.60, 0.90)
at most recent visit	1.34 (1.13, 1.61)
Antiarrhythmic agents	
at baseline	0.58 (0.42, 0.80)
at most recent visit	1.37 (1.11, 1.69)
Diuretics	

at baseline	1.19 (1.00, 1.42)
at most recent visit	0.86 (0.74, 1.01)
Antihypertensives other than diuretics	
at baseline	1.17 (0.92, 1.48)
at most recent visit	0.89 (0.72, 1.10)
<i>Laboratory findings:</i>	
Serum total bilirubin (≥ 0.5 mg/dl)	
at baseline	1.02 (0.92, 1.14)
at most recent visit	0.93 (0.84, 1.03)
Serum total cholesterol (≥ 250 mg/dl)	
at baseline	0.88 (0.78, 0.99)
at most recent visit	1.03 (0.92, 1.16)
Serum triglyceride (≥ 5.0 meq/L)	
at baseline	1.11 (0.99, 1.25)
at most recent visit	0.91 (0.81, 1.02)
Serum uric acid (≥ 7.0 mg/dl)	
at baseline	0.93 (0.82, 1.04)
at most recent visit	1.13 (1.00, 1.28)
Serum alkaline phosphatase (≥ 7.5 K-A units)	
at baseline	1.13 (1.00, 1.27)
at most recent visit	0.93 (0.82, 1.05)
Plasma urea nitrogen (≥ 16 mg/dl)	
at baseline	0.94 (0.84, 1.05)
at most recent visit	1.01 (0.90, 1.12)
Plasma fasting glucose (≥ 100 mg/dl)	
at baseline	0.89 (0.79, 1.01)
at most recent visit	1.01 (0.90, 1.13)
Plasma one-hour glucose, after 75g oral load (≥ 180 mg/dl)	
at baseline	0.91 (0.80, 1.03)
at most recent visit	1.17 (1.04, 1.33)
White blood cell count (≥ 7500 cells/mm ³)	
at baseline	0.99 (0.86, 1.14)
at most recent visit	1.05 (0.91, 1.21)
Absolute neutrophil count (≥ 4500 cells/mm ³)	
at baseline	1.02 (0.89, 1.16)
at most recent visit	1.01 (0.87, 1.15)
Hematocrit ($\geq 46\%$)	
at baseline	1.02 (0.91, 1.14)
at most recent visit	0.90 (0.81, 1.01)
<i>ECG findings:</i>	
Q/QS pattern on antero-lateral, postero-inferior, or antero-septal recording	
at baseline	0.85 (0.74, 0.97)
at most recent visit	1.07 (0.94, 1.22)
ST segment depression on antero-lateral, postero-inferior, or antero-septal recording	
at baseline	1.19 (1.02, 1.39)
at most recent visit	0.99 (0.86, 1.15)
T-wave findings on antero-lateral, postero-inferior, or antero-septal recording	
at baseline	1.03 (0.90, 1.19)
at most recent visit	1.04 (0.90, 1.20)
ST segment elevation on antero-lateral, postero-inferior, or antero-septal recording	
at baseline	0.60 (0.44, 0.82)
at most recent visit	1.33 (1.02, 1.73)
Ventricular conduction defect	
at baseline	0.82 (0.55, 1.23)
at most recent visit	0.90 (0.65, 1.25)
Premature ventricular beats (frequent ventricular ectopic beats)	
at baseline	1.41 (1.05, 1.90)
at most recent visit	1.41 (1.10, 1.80)
Heart rate (≥ 70 beats per minute)	
at baseline	0.85 (0.76, 0.96)
at most recent visit	1.13 (1.01, 1.27)

Table A3. Outcome status and baseline characteristics for censored and uncensored individuals in the updated analyses, Coronary Drug Project.

Covariate	Prevalence among censored individuals, N = 217 (%)	Prevalence among uncensored individuals, N = 2413 (%)
<i>Outcome status</i>		
Death within 5 years	15.2	20.4
<i>Adherence history:</i>		
Low adherence pre-baseline	11.5	6.0
<i>Demographics:</i>		
Age ≥ 55 years at baseline	38.3	43.4
Race (not white)	8.3	6.8
<i>Lifestyle characteristics:</i>		
Cigarette smoking	52.1	36.1
Current habitual level of physical activity in leisure time (light or sedentary)	71.9	69.1
<i>Medical history:</i>		
Risk group (>1 previous MI or MI with complications)	30.4	34.3
Number of prior MIs (≥ 2)	19.4	19.3
Relative body weight (≥ 1.15)	44.7	46.0
Congestive heart failure	12.9	15.9
Angina pectoris	58.1	57.4
Acute coronary insufficiency	18.4	15.8
Intermittent cerebral ischemic attack	3.7	3.4
Intermittent claudication	10.1	8.3
Systolic blood pressure (≥ 130 mmHg)	56.7	50.4
Diastolic blood pressure (≥ 85 mmHg)	46.5	35.4
New York Heart Association functional class (any limitation)	56.2	53.1
Cardiomegaly on chest X-ray	17.5	17.8
<i>Prescription of non-study medications:</i>		
Oral hypoglycemic agents	5.5	5.1
Digitalis	10.1	14.7
Antiarrhythmic agents	2.8	3.5
Diuretics	14.3	13.6
Antihypertensives other than diuretics	8.3	6.7
<i>Laboratory findings:</i>		
Serum total bilirubin (≥ 0.5 mg/dl)	52.5	51.6
Serum total cholesterol (≥ 250 mg/dl)	50.7	45.9
Serum triglyceride (≥ 5.0 meq/L)	50.2	49.4
Serum uric acid (≥ 7.0 mg/dl)	43.3	43.8
Serum alkaline phosphatase (≥ 7.5 K-A units)	46.1	47.8
Plasma urea nitrogen (≥ 16 mg/dl)	47.5	53.1
Plasma fasting glucose (≥ 100 mg/dl)	48.4	40.9
Plasma one-hour glucose, after 75g oral load (≥ 180 mg/dl)	41.0	39.2
White blood cell count (≥ 7500 cells/mm ³)	48.4	44.2
Absolute neutrophil count (≥ 4500 cells/mm ³)	51.2	44.7
Hematocrit ($\geq 46\%$)	60.8	58.1
<i>ECG findings:</i>		
Q/QS pattern on antero-lateral, postero-inferior, or antero-septal recording	64.5	61.3
ST segment depression on antero-lateral, postero-inferior, or antero-septal recording	22.6	24.7
T-wave findings on antero-lateral, postero-inferior, or antero-septal recording	53.0	48.5
ST segment elevation on antero-lateral, postero-inferior, or antero-septal recording	6.0	3.8

Ventricular conduction defect	1.8	4.3
Premature ventricular beats (frequent ventricular ectopic beats)	2.3	2.7
Heart rate (≥ 70 beats per minute)	53.5	43.9

Table A4. Sensitivity analyses for the difference in 5-year mortality for <80% adherence minus ≥80% adherence among 2630 individuals with complete baseline data in the placebo arm, Coronary Drug Project

Sensitivity analysis	5-year mortality difference, % (95% CI)			
	N	Unadjusted	Adjusted for baseline variables	Further adjusted for post-randomization variables
Method to assign % adherence at missed visits % adherence carried forward for 2 missed visits, set to 0 from third consecutive missed visit onwards	2626	6.0 (2.3, 9.7)	3.1 (-0.3, 6.6)	0.2 (-3.7, 4.1)
% adherence carried forward 2 missed visits, set to 80% from third consecutive missed visit onwards.	2626	8.4 (4.4, 12.5)	4.9 (1.1, 8.6)	0.9 (-3.2, 4.9)
% adherence set to cumulative % adherence through last available visit at all subsequent missed visits	2623	8.6 (4.5, 12.7)	5.1 (1.3, 8.8)	1.1 (-3.4, 5.7)
% adherence set to half the adherence at last available visit for all missed visits	2623	7.9 (4.3, 11.5)	4.8 (1.4, 8.1)	1.4 (-2.6, 5.4)
% adherence set to half the adherence at last available visit up to 2 missed visits, censored from third consecutive missed visit onwards	2413	12.5 (8.2, 16.9)	8.2 (4.3, 12.1)	3.6 (-1.0, 8.1)
% adherence set to 0 for 2 missed visits, censored from third consecutive missed visit onwards	2413	11.7 (7.5, 15.8)	3.1 (-0.3, 6.6)	3.2 (-1.2, 7.5)
Original 1980 analysis (missing adherence set to 0 at all visits) with	2627	---	---	7.6 (2.0, 13.1)

adjustment for post-randomization covariates				
Ignore prescribed dosage when calculating adherence No adjustment for censoring	2415	11.6 (5.8, 17.4)	7.6 (2.3, 13.0)	5.8 (0.2, 11.5)
Adjustment for censoring via IP weighting	2385	12.0 (6.3, 17.8)	8.3 (3.1, 13.4)	6.3 (-2.7, 15.2)
Definition of adherence Binary adherence cut-point at 70%	2413	13.5 (6.9, 20.1)	8.2 (1.9, 14.4)	---
Binary adherence cut-point at 90%	2413	-0.2 (-3.4, 3.0)	-2.3 (-5.4, 0.8)	---
Exclude individuals who started after Aug 30, 1969 No adjustment for censoring	2315	11.0 (6.3, 15.6)	6.7 (2.4, 11.0)	2.5 (-2.1, 7.0)
Adjustment for censoring via IP weighting	2289	11.0 (6.3, 15.8)	7.0 (2.6, 11.4)	2.7 (-4.0, 9.3)
Updated analysis, with adjustment for censoring via IP censoring weights	2385	11.2 (6.5, 16.0)	7.3 (3.0, 11.6)	2.5 (-3.8, 8.8)
Updated analysis, with covariates set to median values	2410	---	4.1 (1.1, 7.2)	---

SAS Appendix

This package contains the SAS programs used for the main analyses to replicate the 1980 CDP findings and update these analyses using inverse probability weighting. SAS 9.4 was used for all analyses. The code appendix contains the following programs:

1. Program 1: Data management
This program reads in the CDP dataset and defines the variables for the original and updated analyses. It creates two datasets for the placebo arm. Changing the value of ITR to 3 produces the same datasets for the clofibrate arm.
 - Binary.sas7bdat for crude and baseline adjusted analyses
 - Censwt_ag.sas7bdat for post-randomization adjusted analyses
2. Program 2: Replication of 1980 analysis
This program estimates the crude and baseline adjustment using the original 1980 definition of adherence. The output gives the first 2 columns of row 1 of Table 2 when using the original adherence coding and the second column of row 2 when using the updated adherence coding. It also outputs the results for Table 1.
3. Program 3: Updated analyses with baseline variables
This program estimates the crude and baseline adjustment using updated definition of adherence and outputs columns 1 and 3 of row 2 of Table 2. Changing the adherence variable input allows estimation of the third column of row 1 in Table 2.
4. Program 4: Updated analysis with post-randomization covariates
This program requires the restricted cubic spline macro developed by F.E. Harrell[7, 8] in order to run. It estimates the association between adherence and death using adjustment for post-randomization covariates using inverse probability weighting and outputs final column for row 2 of Table 2.

The datasets for the Coronary Drug Project will be available through application to the National Heart, Lung, and Blood Institute. If you have any questions, comments, or discover an error, please contact Eleanor Murray at emurray@mail.harvard.edu. For the most updated versions of the SAS programs, please visit www.hsph.harvard.edu/causal/.

```

/*****/

/**   Program 1 Data Management *****/
/**This program creates 2 data files for**/
/**analysis of CDP adherence data in*****/
/**placebo arm. Creates: *****/
/*****/
/**Binary.sas7bdat: main dataset with: ***/
/**Old and new adherence definitions*****/
/**dichotomized covariates.*****/
/*****/
/*****/
/**censwt_ag.sas7bdat: long-format *****/
/**dataset for analysis of time-varying*****/
/**adherence*****/
/*****/

```

```
libname cdp "</path/>";
```

```

/*****/
/****Main analysis dataset *****/

```

```

data cdp.binary;
set cdp.allcdp_new;
where itr = 6; /*ITR = 6 is placebo arm. */

```

```
/*For clofibrate arm, use ITR = 3*/
```

```

keep
/*treatment group and administrative variables*/
ID ITR INV1--INV24 INV28 INVCV1--INVCV3 LVA LV5 LV6 LVSS ENTRYDATE ENDDATE DEATHDATE

```

```

/*exposures*/
ICAP1--ICAP26 ICAPQ1 ICAPQ2 IADH1--IADH26
ADHX15 ADH1--ADH52 CAPS1--CAPS26 JADH1-JADH2

```

```

/*outcomes*/
DTH5 ISTAT ITS ITD INVDTH ITDX

```

```

/*covariates from NEJM paper*/
IAGE IRC NMI LMI ZLABB1 ZLABB4
ZLABB5 ZLABB6 ZLABB7 ZLABB8 ZLABB9
ZLABB10 WTBP61 WTBP32 RBW WBC23
WBC1 WBC12 IRK NIHA1
KLIN1 KLIN2 KLIN3 KLIN5 KLIN7

```



```

IMDCT58 IMDCT114 IMDCT86 IMDCT142 IMDCT30
IXRAY1 IXRAY12 ICG1 IJOB34 IECG2
IECG14 IECG26 IECG38 IECG50 IECG62
IECG74 IECG86 IECG98 IECG110 IECG122
IECG134 IECG206 IECG194 IECG278

```

```

ZLAB88 ZLAB89 ZLAB90 ZLAB91--ZLAB116

```

```

/*Longitudinal variables for time-varying adherence analyses*/
NIHA1-NIHA6 ICG1-ICG6 WTBP32-WTBP47 WTBP61-WTBP76
IJOB34-IJOB39 IXRAY1-IXRAY22 WBC1-WBC33 IMDCT30-IMDCT45
IMDCT58-IMDCT73 IMDCT86-IMDCT101 IMDCT114-IMDCT129
IMDCT142-IMDCT157 ICCL1-ICCL15 ZLAB4-ZLAB18 ZLAB91-ZLAB105
ZLAB120-ZLAB134 ZLAB149-ZLAB163 ZLAB178-ZLAB192

```

```

ZLAB207-ZLAB221 ZLAB236-ZLAB250 ZLAB265-ZLAB279
IECG1-IECG283;

```

```

/****Demographic & Baseline-Only Variables****/

```

```

age_bin = .;
if iage ge 55 then age_bin = 1;
else if .< iage <55 then age_bin = 0;

```

```

age_cat = .;
if .< iage <45 then age_cat =0;
else if 45 le iage < 50 then age_cat = 1;
else if 50 le iage < 55 then age_cat = 2;
else if 55 le iage < 60 then age_cat = 3;
else if 60 le iage then age_cat = 4;

```

```

nonwhite = .;
if irc = 1 then nonwhite = 0;
else if irc in (2,3) then nonwhite = 1;

```

```

mi_bin = .;
if NMI ge 2 then mi_bin = 1;
else if .< NMI <2 then mi_bin = 0;

```

```

rbw_bin = .;
if RBW ge 1.15 then rbw_bin = 1;
else if . < RBW < 1.15 then rbw_bin = 0;

```

```

keep age_cat age_bin nonwhite mi_bin rbw_bin;
drop irc nmi rbw;

```

```

/****Time-varying Covariates*****/

```

```
/*For annual variables, code assigns most recent annual visit to subsequent 2 non-annual follow-up visits*/
```

```
/*For any missing data, carry forward up to 2 consecutive missed visits - censor on third consecutive missed visit*/
```

```
/*Where necessary, set invalid response levels to missing*/
```

```
/*New York Heart Association Risk group: 1 v 2 at baseline; 1 v (2,3,4) over follow-up*/
```

```
array NIHA_FV(17) NIHA_FV1 - NIHA_FV17;
```

```
array NIHA(6) NIHA1 - NIHA6;
```

```
array NIHA_bin(6) NIHA_bin1 - NIHA_bin6;
```

```
do i = 1 to 6;
```

```
    if NIHA(i) = 1 then NIHA_bin(i) = 0;
```

```
    else if NIHA(i) = 9 then do;
```

```
        NIHA(i) = . ;
```

```
        NIHA_bin(i) = . ;
```

```
    end;
```

```
    else if NIHA(i) >1 then NIHA_bin(i) = 1;
```

```
end;
```

```
NIHA_FV0 = NIHA_bin1;
```

```
NIHA_FV1 = NIHA_bin1;
```

```
NIHA_FV2 = NIHA_bin1;
```

```
do i = 2 to 6;
```

```
    if NIHA_bin(i) ne . then do;
```

```
        NIHA_FV((i-1)*3) = NIHA_bin(i);
```

```
        NIHA_FV((i-1)*3+1) = NIHA_bin(i);
```

```
        NIHA_FV((i-1)*3+2) = NIHA_bin(i);
```

```
    end;
```

```
    else do;
```

```
        NIHA_FV((i-1)*3) = .;
```

```
        NIHA_FV((i-1)*3+1) = .;
```

```
        NIHA_FV((i-1)*3+2) = .;
```

```
    end;
```

```
end;
```

```
keep NIHA_FV0-NIHA_FV15 NIHA_bin1;
```

```
drop NIHA1-NIHA6 NIHA_FV16-NIHA_FV17;
```

```
/*Clinical history variables: 0=no, 1=suspect or definite; baseline and follow-up*/
```

```
array KLIN1_FV(15) KLIN1_FV1-KLIN1_FV15;
```

```
array KLIN2_FV(15) KLIN2_FV1-KLIN2_FV15;
```

```
array KLIN3_FV(15) KLIN3_FV1-KLIN3_FV15;
```

```
array KLIN5_FV(15) KLIN5_FV1-KLIN5_FV15;
```

```
array KLIN7_FV(15) KLIN7_FV1-KLIN7_FV15;
```

```
array ICCL(75) ICCL1-ICCL15
```

```
            ICCL27-ICCL41
```

```
            ICCL53-ICCL67
```

```
            ICCL157-ICCL171
```

```

        ICCL235-ICCL249;
KLIN1_FV0 = KLIN1;
KLIN2_FV0 = KLIN2;
KLIN3_FV0 = KLIN3;
KLIN5_FV0 = KLIN5;
KLIN7_FV0 = KLIN7;
do i=1 to 15;
    if ICCL(i) ne . then KLIN1_FV(i) = ICCL(i);
    else if i = 1 then KLIN1_FV(i) = KLIN1_FV0;
    else if i = 2 and ICCL(i-1) ne . then KLIN1_FV(i) = KLIN1_FV1;
    else if i > 2 then do;
        if ICCL(i-1) ne . then KLIN1_FV(i) = KLIN1_FV(i-1);
        else if ICCL(i-2) ne . then KLIN1_FV(i) = KLIN1_FV(i-2);
        else if ICCL(i-2) = . then KLIN1_FV(i) = .;
    end;
end;
do i=16 to 30;
    if ICCL(i) ne . then KLIN2_FV(i-15) = ICCL(i);
    else if i = 16 then KLIN2_FV(i-15) = KLIN2_FV0;
    else if i = 17 and ICCL(i-1) ne . then KLIN2_FV(i-15) = KLIN2_FV1;
    else if i > 17 then do;
        if ICCL(i-1) ne . then KLIN2_FV(i-15) = KLIN2_FV(i-16);
        else if ICCL(i-2) ne . then KLIN2_FV(i-15) = KLIN2_FV(i-17);
        else if ICCL(i-2) = . then KLIN2_FV(i-15) = .;
    end;
end;
do i=31 to 45;
    if ICCL(i) ne . then KLIN3_FV(i-30) = ICCL(i);
    else if i = 31 then KLIN3_FV(i-30) = KLIN3_FV0;
    else if i = 32 and ICCL(i-1) ne . then KLIN3_FV(i-30) = KLIN3_FV1;
    else if i > 32 then do;
        if ICCL(i-1) ne . then KLIN3_FV(i-30) = KLIN3_FV(i-31);
        else if ICCL(i-2) ne . then KLIN3_FV(i-30) = KLIN3_FV(i-32);
        else if ICCL(i-2) = . then KLIN3_FV(i-30) = .;
    end;
end;
do i=46 to 60;
    if ICCL(i) ne . then KLIN5_FV(i-45) = ICCL(i);
    else if i = 46 then KLIN5_FV(i-45) = KLIN5_FV0;
    else if i = 47 and ICCL(i-1) ne . then KLIN5_FV(i-45) = KLIN5_FV1;
    else if i > 47 then do;
        if ICCL(i-1) ne . then KLIN5_FV(i-45) = KLIN5_FV(i-46);
        else if ICCL(i-2) ne . then KLIN5_FV(i-45) = KLIN5_FV(i-47);
        else if ICCL(i-2) = . then KLIN5_FV(i-45) = .;
    end;
end;
do i=61 to 75;
    if ICCL(i) ne . then KLIN7_FV(i-60) = ICCL(i);

```

```

    else if i=61                                then KLIN7_FV(i-60) = KLIN7_FV0;
    else if i=62 and ICCL(i-1) ne .              then KLIN7_FV(i-60) = KLIN7_FV1;
    else if i>62 then do;
        if ICCL(i-1) ne .                        then KLIN7_FV(i-60) = KLIN7_FV(i-61);
        else if ICCL(i-2) ne .                  then KLIN7_FV(i-60) = KLIN7_FV(i-62);
        else if ICCL(i-2) = .                   then KLIN7_FV(i-60) = .;
    end;
end;

CHF = .;
if KLIN1 = 2                                then CHF = 0;
if KLIN1 in (1,3) then CHF = 1;
ACI = .;
if KLIN3 = 2                                then ACI = 0;
if KLIN3 in (1, 3)                          then ACI = 1;
AP = .;
if KLIN2 = 2                                then AP = 0;
if KLIN2 in (1,3) then AP = 1;
IC = .;
if KLIN7 = 2                                then IC = 0;
if KLIN7 in (1,3) then IC = 1;
ICIA = .;
if KLIN5 = 2                                then ICIA = 0;
if KLIN5 in (1,3) then ICIA = 1;
keep CHF ACI AP IC ICIA;

array CHF_FV(15) CHF_FV1 - CHF_FV15;
CHF_FV0 = CHF;
do i = 1 to 15;
    if KLIN1_FV(i) = 2 then CHF_FV(i) = 0;
    if KLIN1_FV(i) in (1,3) then CHF_FV(i) = 1;
end;
array ACI_FV(15) ACI_FV1 - ACI_FV15;
ACI_FV0 = ACI;
do i = 1 to 15;
    if KLIN3_FV(i) = 2 then ACI_FV(i) = 0;
    if KLIN3_FV(i) in (1,3) then ACI_FV(i) = 1;
end;
array AP_FV(15) AP_FV1 - AP_FV15;
AP_FV0 = AP;
do i = 1 to 15;
    if KLIN2_FV(i) = 2 then AP_FV(i) = 0;
    if KLIN2_FV(i) in (1,3) then AP_FV(i) = 1;
end;
array IC_FV(15) IC_FV1 - IC_FV15;
IC_FV0 = IC;
do i = 1 to 15;
    if KLIN7_FV(i) = 2 then IC_FV(i) = 0;

```

```

        if KLIN7_FV(i) in (1,3) then IC_FV(i) = 1;
    end;
array ICIA_FV(15) ICIA_FV1 - ICIA_FV15;
ICIA_FV0 = ICIA;
do i = 1 to 15;
    if KLIN5_FV(i) = 2 then ICIA_FV(i) = 0;
    if KLIN5_FV(i) in (1,3) then ICIA_FV(i) = 1;
end;
keep CHF_FV0-CHF_FV15 ACI_FV0-ACI_FV15 AP_FV0-AP_FV15 IC_FV0-IC_FV15 ICIA_FV0-ICIA_FV15;

drop KLIN1_FV0-KLIN1_FV15 KLIN2_FV0-KLIN2_FV15 KLIN3_FV0-KLIN3_FV15 KLIN5_FV0-KLIN5_FV15
KLIN7_FV0-KLIN7_FV15
    ICCL1-ICCL15 ICCL27-ICCL41 ICCL53-ICCL67 ICCL157-ICCL171 ICCL235-ICCL249;

/*Medication use variables: 0 = no, 1 = yes; baseline and follow-up*/
array IMDCT(80) IMDCT30-IMDCT45
                    IMDCT58-IMDCT73
                    IMDCT86-IMDCT101
                    IMDCT114-IMDCT129
                    IMDCT142-IMDCT157;
array IMDCT30_FV(17) IMDCT30_FV1-IMDCT30_FV17;
array IMDCT58_FV(17) IMDCT58_FV1-IMDCT58_FV17;
array IMDCT86_FV(17) IMDCT86_FV1-IMDCT86_FV17;
array IMDCT114_FV(17) IMDCT114_FV1-IMDCT114_FV17;
array IMDCT142_FV(17) IMDCT142_FV1-IMDCT142_FV17;
do i = 1 to 80;
    if IMDCT(i) = 0 then IMDCT(i) = .;
end;
IMDCT30_FV0 = IMDCT30;
IMDCT58_FV0 = IMDCT58;
IMDCT86_FV0 = IMDCT86;
IMDCT114_FV0 = IMDCT114;
IMDCT142_FV0 = IMDCT142;
do i = 2 to 16;
    if IMDCT(i) ne . then IMDCT30_FV(i-1) = IMDCT(i);
    else if i = 2 then IMDCT30_FV(i-1) = IMDCT30_FV0;
    else if i = 3 and IMDCT(i-1) ne . then IMDCT30_FV(i-1) = IMDCT30_FV1;
    else if i > 3 then do;
        if IMDCT(i-1) ne . then IMDCT30_FV(i-1) = IMDCT30_FV(i-2);
        else if IMDCT(i-2) ne . then IMDCT30_FV(i-1) = IMDCT30_FV(i-3);
        else if IMDCT(i-2) = . then IMDCT30_FV(i-1) = .;
    end;
end;
do i = 18 to 32;
    if IMDCT(i) ne . then IMDCT58_FV(i-17) = IMDCT(i);
    else if i = 18 then IMDCT58_FV(i-17) = IMDCT58_FV0;
    else if i = 19 and IMDCT(i-1) ne . then IMDCT58_FV(i-17) = IMDCT58_FV1;
    else if i > 19 then do;

```

```

        if IMDCT(i-1) ne .          then IMDCT58_FV(i-17) = IMDCT58_FV(i-18);
        else if IMDCT(i-2) ne .    then IMDCT58_FV(i-17) = IMDCT58_FV(i-19);
        else if IMDCT(i-2) = .      then IMDCT58_FV(i-17) = .;
    end;
end;
do i=34 to 48;
    if IMDCT(i) ne .          then IMDCT86_FV(i-33) = IMDCT(i);
    else if i = 34             then IMDCT86_FV(i-33) = IMDCT86_FV0;
    else if i = 35 and IMDCT(i-1) ne . then IMDCT86_FV(i-33) = IMDCT86_FV1;
    else if i > 35 then do;
        if IMDCT(i-1) ne .          then IMDCT86_FV(i-33) = IMDCT86_FV(i-34);
        else if IMDCT(i-2) ne .    then IMDCT86_FV(i-33) = IMDCT86_FV(i-35);
        else if IMDCT(i-2) = .      then IMDCT86_FV(i-33) = .;
    end;
end;
do i=50 to 64;
    if IMDCT(i) ne .          then IMDCT114_FV(i-49) = IMDCT(i);
    else if i = 50             then IMDCT114_FV(i-49) = IMDCT114_FV0;
    else if i = 51 and IMDCT(i-1) ne . then IMDCT114_FV(i-49) = IMDCT114_FV1;
    else if i > 51 then do;
        if IMDCT(i-1) ne .          then IMDCT114_FV(i-49) = IMDCT114_FV(i-50);
        else if IMDCT(i-2) ne .    then IMDCT114_FV(i-49) = IMDCT114_FV(i-51);
        else if IMDCT(i-2) = .      then IMDCT114_FV(i-49) = .;
    end;
end;
do i=66 to 80;
    if IMDCT(i) ne .          then IMDCT142_FV(i-65) = IMDCT(i);
    else if i = 66             then IMDCT142_FV(i-65) = IMDCT142_FV0;
    else if i = 67 and IMDCT(i-1) ne . then IMDCT142_FV(i-65) = IMDCT142_FV1;
    else if i > 67 then do;
        if IMDCT(i-1) ne .          then IMDCT142_FV(i-65) = IMDCT142_FV(i-66);
        else if IMDCT(i-2) ne .    then IMDCT142_FV(i-65) = IMDCT142_FV(i-67);
        else if IMDCT(i-2) = .      then IMDCT142_FV(i-65) = .;
    end;
end;

```

```

DIG = .;
if IMDCT58 = 2 then DIG = 0;
if IMDCT58 = 1 then DIG = 1;
DIUR = .;
if IMDCT114 = 2 then DIUR = 0;
if IMDCT114 = 1 then DIUR = 1;
AntiArr = .;
if IMDCT86 = 2 then AntiArr = 0;
if IMDCT86 = 1 then AntiArr = 1;
AntiHyp = .;

```

```

if IMDCT142 = 2 then AntiHyp = 0;
if IMDCT142 = 1 then AntiHyp = 1;
OralHyp =.;
if IMDCT30 = 2 then OralHyp = 0;
if IMDCT30 = 1 then OralHyp = 1;
Keep DIG DIUR AntiArr AntiHyp OralHyp;

array DIG_FV(15) DIG_FV1 - DIG_FV15;
DIG_FV0 = DIG;
do i = 1 to 15;
    if IMDCT58_FV(i) = 2 then DIG_FV(i) = 0;
    if IMDCT58_FV(i) = 1 then DIG_FV(i) = 1;
end;
array DIUR_FV(15) DIUR_FV1 - DIUR_FV15;
DIUR_FV0 = DIUR;
do i = 1 to 15;
    if IMDCT114_FV(i) = 2 then DIUR_FV(i) = 0;
    if IMDCT114_FV(i) = 1 then DIUR_FV(i) = 1;
end;
array AntiArr_FV(15) AntiArr_FV1 - AntiArr_FV15;
AntiArr_FV0 = AntiArr;
do i = 1 to 15;
    if IMDCT86_FV(i) = 2 then AntiArr_FV(i) = 0;
    if IMDCT86_FV(i) = 1 then AntiArr_FV(i) = 1;
end;
array AntiHyp_FV(15) AntiHyp_FV1 - AntiHyp_FV15;
AntiHyp_FV0 = AntiHyp;
do i = 1 to 15;
    if IMDCT142_FV(i) = 2 then AntiHyp_FV(i) = 0;
    if IMDCT142_FV(i) = 1 then AntiHyp_FV(i) = 1;
end;
array OralHyp_FV(15) OralHyp_FV1 - OralHyp_FV15;
OralHyp_FV0 = OralHyp;
do i = 1 to 15;
    if IMDCT30_FV(i) = 2 then OralHyp_FV(i) = 0;
    if IMDCT30_FV(i) = 1 then OralHyp_FV(i) = 1;
end;
keep DIG_FV0-DIG_FV15 DIUR_FV0-DIUR_FV15 AntiArr_FV0-AntiArr_FV15
AntiHyp_FV0-AntiHyp_FV15 OralHyp_FV0-OralHyp_FV15;

drop IMDCT30-IMDCT45
                                IMDCT58-IMDCT73
                                IMDCT86-IMDCT101
                                IMDCT114-IMDCT129
                                IMDCT142-IMDCT157;

/*Cardiomegaly on x-ray: 0 = no, 1 = probable or definite; baseline and follow-up*/
array IXRAY1_FV(17) IXRAY1_FV1-IXRAY1_FV17;

```

```

array IXRAY2_FV(17) IXRAY2_FV1-IXRAY2_FV17;
array IXRAY(22) IXRAY1 - IXRAY22;
do i = 1 to 11;
    if IXRAY(i) = 3 then IXRAY(i)=.;
end;
IXRAY1_FV0 = IXRAY1;
IXRAY1_FV1 = IXRAY1;
IXRAY1_FV2 = IXRAY1;
do i = 2 to 6;
    if IXRAY(i) ne . then do;
        IXRAY1_FV((i-1)*3) = IXRAY(i);
        IXRAY1_FV((i-1)*3+1) = IXRAY(i);
        IXRAY1_FV((i-1)*3+2) = IXRAY(i);
    end;
    else do;
        IXRAY1_FV((i-1)*3) = .;
        IXRAY1_FV((i-1)*3+1) = .;
        IXRAY1_FV((i-1)*3+2) = .;
    end;
end;
IXRAY2_FV0 = IXRAY12;
IXRAY2_FV1 = IXRAY12;
IXRAY2_FV2 = IXRAY12;
do i = 13 to 17;
    if IXRAY(i) ne . then do;
        IXRAY2_FV((i-12)*3) = IXRAY(i);
        IXRAY2_FV((i-12)*3+1) = IXRAY(i);
        IXRAY2_FV((i-12)*3+2) = IXRAY(i);
    end;
    else do;
        IXRAY2_FV((i-12)*3) = .;
        IXRAY2_FV((i-12)*3+1) = .;
        IXRAY2_FV((i-12)*3+2) = .;
    end;
end;

CardioM =.;
if IXRAY1 = 1 or IXRAY12 = 1 then CardioM = 0;
if IXRAY12 in (2,3) then CardioM = 1;
keep CardioM;

array CardioM_FV(15) CardioM_FV1 - CardioM_FV15;
CardioM_FV0 = CardioM;
do i = 1 to 15;
    if IXRAY1_FV(i) = 1 or IXRAY2_FV(i) = 1 then CardioM_FV(i) = 0;
    if IXRAY2_FV(i) in (2,3) then CardioM_FV(i) = 1;
end;
keep CardioM_FV0 - CardioM_FV15;

```



```
drop IXRAY1 - IXRAY22;
```

```
/*ECG findings: categorical; baseline and follow-up, scheduled ECGs only*/
```

```
array IECG(283) IECG1-IECG283;
array IECG2_FV(17) IECG2_FV1 - IECG2_FV17;
array IECG14_FV(17) IECG14_FV1 - IECG14_FV17;
array IECG26_FV(17) IECG26_FV1 - IECG26_FV17;
array IECG38_FV(17) IECG38_FV1 - IECG38_FV17;
array IECG50_FV(17) IECG50_FV1 - IECG50_FV17;
array IECG62_FV(17) IECG62_FV1 - IECG62_FV17;
array IECG74_FV(17) IECG74_FV1 - IECG74_FV17;
array IECG86_FV(17) IECG86_FV1 - IECG86_FV17;
array IECG98_FV(17) IECG98_FV1 - IECG98_FV17;
array IECG110_FV(17) IECG110_FV1 - IECG110_FV17;
array IECG122_FV(17) IECG122_FV1 - IECG122_FV17;
array IECG134_FV(17) IECG134_FV1 - IECG134_FV17;
array IECG206_FV(17) IECG206_FV1 - IECG206_FV17;
array IECG194_FV(17) IECG194_FV1 - IECG194_FV17;
array IECG278_FV(17) IECG278_FV1 - IECG278_FV17;
do i=1 to 36;
    if IECG(i) in (1, 10, 20, 37, 99) then IECG(i) = .;
end;
do i=37 to 72;
    if IECG(i) in (5, 8, 9) then IECG(i) = .;
end;
do i=73 to 108;
    if IECG(i) in (6, 7, 9) then IECG(i) = .;
end;
do i=109 to 144;
    if IECG(i) in (3, 9) then IECG(i) = .;
end;
do i=206 to 211;
    if IECG(i) in (6, 9) then IECG(i) = .;
end;
do i=194 to 199;
    if IECG(i) = 9 then IECG(i) = .;
end;
do i=278 to 283;
    if IECG(i) in (490, 999) then IECG(i) = .;
end;
IECG2_FV0 = IECG2;
IECG14_FV0 = IECG14;
IECG26_FV0 = IECG26;
IECG38_FV0 = IECG38;
IECG50_FV0 = IECG50;
IECG62_FV0 = IECG62;
IECG74_FV0 = IECG74;
```

```

IECG86_FV0 = IECG86;
IECG98_FV0 = IECG98;
IECG110_FV0 = IECG110;
IECG122_FV0 = IECG122;
IECG134_FV0 = IECG134;
IECG206_FV0 = IECG206;
IECG194_FV0 = IECG194;
IECG278_FV0 = IECG278;
IECG2_FV1 = IECG2;
IECG14_FV1 = IECG14;
IECG26_FV1 = IECG26;
IECG38_FV1 = IECG38;
IECG50_FV1 = IECG50;
IECG62_FV1 = IECG62;
IECG74_FV1 = IECG74;
IECG86_FV1 = IECG86;
IECG98_FV1 = IECG98;
IECG110_FV1 = IECG110;
IECG122_FV1 = IECG122;
IECG134_FV1 = IECG134;
IECG206_FV1 = IECG206;
IECG194_FV1 = IECG194;
IECG278_FV1 = IECG278;
IECG2_FV2 = IECG2;
IECG14_FV2 = IECG14;
IECG26_FV2 = IECG26;
IECG38_FV2 = IECG38;
IECG50_FV2 = IECG50;
IECG62_FV2 = IECG62;
IECG74_FV2 = IECG74;
IECG86_FV2 = IECG86;
IECG98_FV2 = IECG98;
IECG110_FV2 = IECG110;
IECG122_FV2 = IECG122;
IECG134_FV2 = IECG134;
IECG206_FV2 = IECG206;
IECG194_FV2 = IECG194;
IECG278_FV2 = IECG278;

```

```
do i = 3 to 7;
```

```
    if IECG(i) ne . then do;
```

```
        IECG2_FV((i-2)*3) = IECG(i);
```

```
        IECG2_FV((i-2)*3+1) = IECG(i);
```

```
        IECG2_FV((i-2)*3+2) = IECG(i);
```

```
    end;
```

```
    else do;
```

```
        IECG2_FV((i-2)*3) = .;
```

```
        IECG2_FV((i-2)*3+1) = .;
```

```

        IECG2_FV((i-2)*3+2) = .;
    end;

end;

do i = 15 to 19;
    if IECG(i) ne . then do;
        IECG14_FV((i-14)*3) = IECG(i);
        IECG14_FV((i-14)*3+1) = IECG(i);
        IECG14_FV((i-14)*3+2) = IECG(i);
    end;
    else do;
        IECG14_FV((i-14)*3) = .;
        IECG14_FV((i-14)*3+1) = .;
        IECG14_FV((i-14)*3+2) = .;
    end;
end;

do i = 27 to 31;
    if IECG(i) ne . then do;
        IECG26_FV((i-26)*3) = IECG(i);
        IECG26_FV((i-26)*3+1) = IECG(i);
        IECG26_FV((i-26)*3+2) = IECG(i);
    end;
    else do;
        IECG26_FV((i-26)*3) = .;
        IECG26_FV((i-26)*3+1) = .;
        IECG26_FV((i-26)*3+2) = .;
    end;
end;

do i = 39 to 43;
    if IECG(i) ne . then do;
        IECG38_FV((i-38)*3) = IECG(i);
        IECG38_FV((i-38)*3+1) = IECG(i);
        IECG38_FV((i-38)*3+2) = IECG(i);
    end;
    else do;
        IECG38_FV((i-38)*3) = .;
        IECG38_FV((i-38)*3+1) = .;
        IECG38_FV((i-38)*3+2) = .;
    end;
end;

do i = 51 to 55;
    if IECG(i) ne . then do;
        IECG50_FV((i-50)*3) = IECG(i);
        IECG50_FV((i-50)*3+1) = IECG(i);
        IECG50_FV((i-50)*3+2) = IECG(i);
    end;
end;

```

```

        end;
    else do;
        IECG50_FV((i-50)*3) = .;
        IECG50_FV((i-50)*3+1) = .;
        IECG50_FV((i-50)*3+2) = .;
    end;
end;

do i = 63 to 67;
    if IECG(i) ne . then do;
        IECG62_FV((i-62)*3) = IECG(i);
        IECG62_FV((i-62)*3+1) = IECG(i);
        IECG62_FV((i-62)*3+2) = IECG(i);
    end;
    else do;
        IECG62_FV((i-62)*3) = .;
        IECG62_FV((i-62)*3+1) = .;
        IECG62_FV((i-62)*3+2) = .;
    end;
end;

do i = 75 to 79;
    if IECG(i) ne . then do;
        IECG74_FV((i-74)*3) = IECG(i);
        IECG74_FV((i-74)*3+1) = IECG(i);
        IECG74_FV((i-74)*3+2) = IECG(i);
    end;
    else do;
        IECG74_FV((i-74)*3) = .;
        IECG74_FV((i-74)*3+1) = .;
        IECG74_FV((i-74)*3+2) = .;
    end;
end;

do i = 87 to 91;
    if IECG(i) ne . then do;
        IECG86_FV((i-86)*3) = IECG(i);
        IECG86_FV((i-86)*3+1) = IECG(i);
        IECG86_FV((i-86)*3+2) = IECG(i);
    end;
    else do;
        IECG86_FV((i-86)*3) = .;
        IECG86_FV((i-86)*3+1) = .;
        IECG86_FV((i-86)*3+2) = .;
    end;
end;

do i = 99 to 103;

```

```

    if IECG(i) ne . then do;
        IECG98_FV((i-98)*3) = IECG(i);
        IECG98_FV((i-98)*3+1) = IECG(i);
        IECG98_FV((i-98)*3+2) = IECG(i);
    end;
    else do;
        IECG98_FV((i-98)*3) = .;
        IECG98_FV((i-98)*3+1) = .;
        IECG98_FV((i-98)*3+2) = .;
    end;
end;

do i = 111 to 115;
    if IECG(i) ne . then do;
        IECG110_FV((i-110)*3) = IECG(i);
        IECG110_FV((i-110)*3+1) = IECG(i);
        IECG110_FV((i-110)*3+2) = IECG(i);
    end;
    else do;
        IECG110_FV((i-110)*3) = .;
        IECG110_FV((i-110)*3+1) = .;
        IECG110_FV((i-110)*3+2) = .;
    end;
end;

do i = 123 to 127;
    if IECG(i) ne . then do;
        IECG122_FV((i-122)*3) = IECG(i);
        IECG122_FV((i-122)*3+1) = IECG(i);
        IECG122_FV((i-122)*3+2) = IECG(i);
    end;
    else do;
        IECG122_FV((i-122)*3) = .;
        IECG122_FV((i-122)*3+1) = .;
        IECG122_FV((i-122)*3+2) = .;
    end;
end;

do i = 135 to 139;
    if IECG(i) ne . then do;
        IECG134_FV((i-134)*3) = IECG(i);
        IECG134_FV((i-134)*3+1) = IECG(i);
        IECG134_FV((i-134)*3+2) = IECG(i);
    end;
    else do;
        IECG134_FV((i-134)*3) = .;
        IECG134_FV((i-134)*3+1) = .;
        IECG134_FV((i-134)*3+2) = .;
    end;
end;

```

```

        end;
    end;

    do i = 207 to 211;
        if IECG(i) ne . then do;
            IECG206_FV((i-206)*3) = IECG(i);
            IECG206_FV((i-206)*3+1) = IECG(i);
            IECG206_FV((i-206)*3+2) = IECG(i);
        end;
        else do;
            IECG206_FV((i-206)*3) = .;
            IECG206_FV((i-206)*3+1) = .;
            IECG206_FV((i-206)*3+2) = .;
        end;
    end;

    do i = 195 to 199;
        if IECG(i) ne . then do;
            IECG194_FV((i-194)*3) = IECG(i);
            IECG194_FV((i-194)*3+1) = IECG(i);
            IECG194_FV((i-194)*3+2) = IECG(i);
        end;
        else do;
            IECG194_FV((i-194)*3) = .;
            IECG194_FV((i-194)*3+1) = .;
            IECG194_FV((i-194)*3+2) = .;
        end;
    end;

    do i = 279 to 283;
        if IECG(i) ne . then do;
            IECG278_FV((i-278)*3) = IECG(i);
            IECG278_FV((i-278)*3+1) = IECG(i);
            IECG278_FV((i-278)*3+2) = IECG(i);
        end;
        else do;
            IECG278_FV((i-278)*3) = .;
            IECG278_FV((i-278)*3+1) = .;
            IECG278_FV((i-278)*3+2) = .;
        end;
    end;

    QQS = .;
    if IECG2 = 0 and IECG14 = 0 and IECG26 = 0 then QQS = 0;
    if IECG2 in (28,31,32,33,34,35,36) or IECG14 in (28,31,32,33,34,35,36)
        or IECG26 in (28,31,32,33,34,35,36) then QQS = 1;
    if IECG2 in (21,22,23,24,25,26,27) or IECG14 in (21,22,23,24,25,26,27)
        or IECG26 in (21,22,23,24,25,26,27) then QQS = 2;

```

```

if IECG2 in (11,12,13,14,15,16,17,18) or IECG14 in (11,12,13,14,15,16,17,18)
    or IECG26 in (11,12,13,14,15,16,17,18) then QQS = 3;
array QQS_FV(15) QQS_FV1 - QQS_FV15;
QQS_FV0 = QQS;
do i = 1 to 15;
    if IECG2_FV(i) = 0 and IECG14_FV(i) = 0 and IECG26_FV(i) = 0 then
QQS_FV(i) = 0;
        if IECG2_FV(i) in (28,31,32,33,34,35,36) or IECG14_FV(i) in (28,31,32,33,34,35,36)
            or IECG26_FV(i) in (28,31,32,33,34,35,36) then
QQS_FV(i) = 1;
            if IECG2_FV(i) in (21,22,23,24,25,26,27) or IECG14_FV(i) in (21,22,23,24,25,26,27)
                or IECG26_FV(i) in (21,22,23,24,25,26,27) then
QQS_FV(i) = 2;
                if IECG2_FV(i) in (11,12,13,14,15,16,17,18) or IECG14_FV(i) in (11,12,13,14,15,16,17,18)
                    or IECG26_FV(i) in (11,12,13,14,15,16,17,18) then
QQS_FV(i) = 3;
                    end;
end;

```

```

AnyQQS = .;
if QQS ge 1 then AnyQQS = 1;
if QQS = 0 then AnyQQS = 0;
array AnyQQS_FV(15) AnyQQS_FV1 - AnyQQS_FV15;
AnyQQS_FV0 = AnyQQS;
do i=1 to 15;
    if QQS_FV(i) ge 1 then AnyQQS_FV(i) = 1;
    if QQS_FV(i) = 0 then AnyQQS_FV(i) = 0;
end;

```

```

STDep = .;
if IECG38 = 0 and IECG50 = 0 and IECG62 = 0 then STDep = 0;
else if IECG38 = 4 or IECG50 = 4 or IECG62 = 4 then STDep = 1;
else if IECG38 = 3 or IECG50 = 3 or IECG62 = 3 then STDep = 2;
else if IECG38 = 2 or IECG50 = 2 or IECG62 = 2 then STDep = 3;
else if IECG38 = 1 or IECG50 = 1 or IECG62 = 1 then STDep = 4;
array STDep_FV(15) STDep_FV1 - STDep_FV15;
STDep_FV0 = STDep;
do i = 1 to 15;
    if IECG38_FV(i) = 0 and IECG50_FV(i) = 0 and IECG62_FV(i) = 0 then STDep_FV(i) = 0;
    else if IECG38_FV(i) = 4 or IECG50_FV(i) = 4 or IECG62_FV(i) = 4 then STDep_FV(i) = 1;
    else if IECG38_FV(i) = 3 or IECG50_FV(i) = 3 or IECG62_FV(i) = 3 then STDep_FV(i) = 2;
    else if IECG38_FV(i) = 2 or IECG50_FV(i) = 2 or IECG62_FV(i) = 2 then STDep_FV(i) = 3;
    else if IECG38_FV(i) = 1 or IECG50_FV(i) = 1 or IECG62_FV(i) = 1 then STDep_FV(i) = 4;
end;

```

```

AnySTDep = .;
if STDep = 0 then AnySTDep = 0;
if STDep > 0 then AnySTDep = 1;
array AnySTDep_FV(15) AnySTDep_FV1 - AnySTDep_FV15;

```

```
AnySTDep_FV0 = AnySTDep;
```

```
do i=1 to 15;
```

```
    if STDep_FV(i) = 0      then AnySTDep_FV(i) = 0;
```

```
    if STDep_FV(i) >0      then AnySTDep_FV(i) = 1;
```

```
end;
```

```
TWave = .;
```

```
if IECG74 = 0 and IECG86 = 0 and IECG98 = 0      then TWave = 0;
```

```
else if IECG74 = 4 or IECG86 = 4 or IECG98 = 4    then TWave = 1;
```

```
else if IECG74 = 3 or IECG86 = 3 or IECG98 = 3    then TWave = 2;
```

```
else if IECG74 = 2 or IECG86 = 2 or IECG98 = 2    then TWave = 3;
```

```
else if IECG74 = 1 or IECG86 = 1 or IECG98 = 1    then TWave = 4;
```

```
array TWave_FV(15) TWave_FV1 - TWave_FV15;
```

```
TWave_FV0 = TWave;
```

```
do i = 1 to 15;
```

```
    if IECG74_FV(i) = 0 and IECG86_FV(i) = 0 and IECG98_FV(i) = 0      then TWave_FV(i) = 0;
```

```
    else if IECG74_FV(i) = 4 or IECG86_FV(i) = 4 or IECG98_FV(i) = 4    then TWave_FV(i) = 1;
```

```
    else if IECG74_FV(i) = 3 or IECG86_FV(i) = 3 or IECG98_FV(i) = 3    then TWave_FV(i) = 2;
```

```
    else if IECG74_FV(i) = 2 or IECG86_FV(i) = 2 or IECG98_FV(i) = 2    then TWave_FV(i) = 3;
```

```
    else if IECG74_FV(i) = 1 or IECG86_FV(i) = 1 or IECG98_FV(i) = 1    then TWave_FV(i) = 4;
```

```
end;
```

```
AnyTwave=.
```

```
if Twave = 0      then AnyTwave = 0;
```

```
if Twave >0      then AnyTwave = 1;
```

```
array AnyTwave_FV(15) AnyTwave_FV1 - AnyTwave_FV15;
```

```
AnyTwave_FV0 = AnyTwave;
```

```
do i=1 to 15;
```

```
    if Twave_FV(i) = 0      then AnyTwave_FV(i) = 0;
```

```
    if Twave_FV(i) >0      then AnyTwave_FV(i) = 1;
```

```
end;
```

```
STelev = .;
```

```
if IECG110 = 0 and IECG122 = 0 and IECG134 = 0 then STElev = 0;
```

```
if IECG110 = 2 or IECG122 = 2 or IECG134 = 2    then STElev = 1;
```

```
array STElev_FV(15) STElev_FV1 - STElev_FV15;
```

```
STelev_FV0 = STElev;
```

```
do i = 1 to 15;
```

```
    if IECG110_FV(i) = 0 and IECG122_FV(i) = 0 and IECG134_FV(i) = 0      then STElev_FV(i) = 0;
```

```
    if IECG110_FV(i) = 2 or IECG122_FV(i) = 2 or IECG134_FV(i) = 2      then STElev_FV(i) = 1;
```

```
end;
```

```
FVEB = .;
```

```
if IECG206 in (0, 1, 3, 4) then FVEB = 0;
```

```
if IECG206 = 2 then FVEB = 1;
```

```
array FVEB_FV(15) FVEB_FV1 - FVEB_FV15;
```

```
FVEB_FV0 = FVEB;
```

```
do i = 1 to 15;
```



```

    if IECG206_FV(i) in (0, 1, 3, 4) then FVEB_FV(i) = 0;
    if IECG206_FV(i) = 2 then FVEB_FV(i) = 1;
    end;

VCD = .;
if IECG194 in (0,3,5) then VCD = 0;
if IECG194 in (1,2,4) then VCD = 1;
array VCD_FV(15) VCD_FV1 - VCD_FV15;
VCD_FV0 = VCD;
do i = 1 to 15;
    if IECG194_FV(i) in (0,3,5) then VCD_FV(i) = 0;
    if IECG194_FV(i) in (1,2,4) then VCD_FV(i) = 1;
    end;

HiHeart = .;
if IECG278 ge 70 then HiHeart = 1;
if IECG278 <70 then HiHeart = 0;
array HiHeart_FV(15) HiHeart_FV1-HiHeart_FV15;
HiHeart_FV0 = HiHeart;
do i=1 to 15;
    if IECG278_FV(i) ge 70 then HiHeart_FV(i) = 1;
    if IECG278_FV(i) <70 then HiHeart_FV(i) = 0;
    end;

keep QQS QQS_FV0-QQS_FV15 AnyQQS AnyQQS_FV0-AnyQQS_FV15
    STDep STDep_FV0-STDep_FV15 AnySTDep AnySTDep_FV0-AnySTDep_FV15
    TWave TWave_FV0-TWave_FV15 AnyTWave AnyTWave_FV0-AnyTWave_FV15
    STElev STElev_FV0-STElev_FV15
    FVEB FVEB_FV0-FVEB_FV15
    VCD VCD_FV0-VCD_FV15
    HiHeart HiHeart_FV0-HiHeart_FV15;

drop IECG1-IECG283;

/* Blood pressure*/
array SysBP_FV(15) SysBP_FV1-SysBP_FV15;
array DiasBP_FV(15) DiasBP_FV1-DiasBP_FV15;
array WTBP(32) WTBP32 - WTBP47 WTBP61-WTBP76;
SysBP_FV0 = WTBP32;
do i=2 to 16;
    if WTBP(i) ne . then SysBP_FV(i-1) = WTBP(i);
    else if i = 2 then SysBP_FV(i-1) = SysBP_FV0;
    else if i = 3 and WTBP(i-1) ne . then SysBP_FV(i-1) = SysBP_FV1;
    else if i > 3 then do;
        if WTBP(i-1) ne . then SysBP_FV(i-1) = SysBP_FV(i-2);
        else if WTBP(i-2) ne . then SysBP_FV(i-1) = SysBP_FV(i-3);
        else if WTBP(i-2) = . then SysBP_FV(i-1) = .;
    end;
end;

```

```

        end;
    end;
    DiasBP_FV0 = WTBP61;
    do i=18 to 32;
        if WTBP(i) ne . then DiasBP_FV(i-17) = WTBP(i);
        else if i = 18 then DiasBP_FV(i-17) = DiasBP_FV0;
        else if i = 19 and WTBP(i-1) ne . then DiasBP_FV(i-17) = DiasBP_FV1;
        else if i > 19 then do;
            if WTBP(i-1) ne . then DiasBP_FV(i-17) = DiasBP_FV(i-18);
            else if WTBP(i-2) ne . then DiasBP_FV(i-17) = DiasBP_FV(i-19);
            else if WTBP(i-2) = . then DiasBP_FV(i-17) = .;
        end;
    end;
end;

```

```

array HiSysBP_a(15) HiSysBP_FV1 - HiSysBP_FV15;
array HiDiasBP_a(15) HiDiasBP_FV1 - HiDiasBP_FV15;
if . < SysBP_FV0 < 130 then do;
    HiSysBP = 0;
    HiSysBP_FV0 = 0;
end;
else if SysBP_FV0 ge 130 then do;
    HiSysBP =1;
    HiSysBP_FV0 = 1;
end;
do i = 1 to 15;
    if . < SysBP_FV(i) < 130 then HiSysBP_a(i) = 0;
    else if SysBP_FV(i) ge 130 then HiSysBP_a(i) =1;
end;

if . < DiasBP_FV0 < 85 then do;
    HiDiasBP = 0;
    HiDiasBP_FV0 = 0;
end;
else if DiasBP_FV0 ge 85 then do;
    HiDiasBP =1;
    HiDiasBP_FV0 = 1;
end;
do i = 1 to 15;
    if . < DiasBP_FV(i) < 85 then HiDiasBP_a(i) = 0;
    else if DiasBP_FV(i) ge 85 then HiDiasBP_a(i) =1;
end;

keep HiSysBP HiSysBP_FV0 - HiSysBP_FV15
HiDiasBP HiDiasBP_FV0 - HiDiasBP_FV15;

drop WTBP32 - WTBP47 WTBP61-WTBP76;

```

```

/*blood work labs*/
array ZLABB1_FV(15) ZLABB1_FV1-ZLABB1_FV15;
array ZLABB4_FV(15) ZLABB4_FV1-ZLABB4_FV15;
array ZLABB5_FV(15) ZLABB5_FV1-ZLABB5_FV15;
array ZLABB6_FV(15) ZLABB6_FV1-ZLABB6_FV15;
array ZLABB7_FV(15) ZLABB7_FV1-ZLABB7_FV15;
array ZLABB8_FV(15) ZLABB8_FV1-ZLABB8_FV15;
array ZLABB9_FV(15) ZLABB9_FV1-ZLABB9_FV15;
array ZLABB10_FV(15) ZLABB10_FV1-ZLABB10_FV15;
array ZLAB(120) ZLAB4-ZLAB18
                                ZLAB91-ZLAB105
                                ZLAB120-ZLAB134
                                ZLAB149-ZLAB163
                                ZLAB178-ZLAB192
                                ZLAB207-ZLAB221
                                ZLAB236-ZLAB250
                                ZLAB265-ZLAB279;

ZLABB1_FV0 = ZLABB1;
ZLABB4_FV0 = ZLABB4;
ZLABB5_FV0 = ZLABB5;
ZLABB6_FV0 = ZLABB6;
ZLABB7_FV0 = ZLABB7;
ZLABB8_FV0 = ZLABB8;
ZLABB9_FV0 = ZLABB9;
ZLABB10_FV0 = ZLABB10;
do i=1 to 15;
    if ZLAB(i) ne .                then ZLABB1_FV(i) = ZLAB(i);
    else if i = 1                  then ZLABB1_FV(i) = ZLABB1_FV0;
    else if i = 2 and ZLAB(i-1) ne . then ZLABB1_FV(i) = ZLABB1_FV1;
    else if i > 2 then do;
        if ZLAB(i-1) ne .          then ZLABB1_FV(i) = ZLABB1_FV(i-1);
        else if ZLAB(i-2) ne .      then ZLABB1_FV(i) = ZLABB1_FV(i-2);
        else if ZLAB(i-2) = .       then ZLABB1_FV(i) = .;
    end;
end;
do i=16 to 30;
    if ZLAB(i) ne .                then ZLABB4_FV(i-15) = ZLAB(i);
    else if i = 16                  then ZLABB4_FV(i-15) = ZLABB4_FV0;
    else if i = 17 and ZLAB(i-1) ne . then ZLABB4_FV(i-15) = ZLABB4_FV1;
    else if i > 17 then do;
        if ZLAB(i-1) ne .          then ZLABB4_FV(i-15) = ZLABB4_FV(i-16);
        else if ZLAB(i-2) ne .      then ZLABB4_FV(i-15) = ZLABB4_FV(i-17);
        else if ZLAB(i-2) = .       then ZLABB4_FV(i-15) = .;
    end;
end;
do i=31 to 45;
    if ZLAB(i) ne .                then ZLABB5_FV(i-30) = ZLAB(i);

```

```

else if i = 31                                then ZLABB5_FV(i-30) = ZLABB5_FV0;
else if i = 32 and ZLAB(i-1) ne . then ZLABB5_FV(i-30) = ZLABB5_FV1;
else if i > 32 then do;
    if ZLAB(i-1) ne .                        then ZLABB5_FV(i-30) = ZLABB5_FV(i-31);
    else if ZLAB(i-2) ne .                    then ZLABB5_FV(i-30) = ZLABB5_FV(i-32);
    else if ZLAB(i-2) = .                      then ZLABB5_FV(i-30) = .;
end;
end;
do i=46 to 60;
    if ZLAB(i) ne .                        then ZLABB6_FV(i-45) = ZLAB(i);
    else if i = 46                          then ZLABB6_FV(i-45) = ZLABB6_FV0;
    else if i = 47 and ZLAB(i-1) ne . then ZLABB6_FV(i-45) = ZLABB6_FV1;
    else if i > 47 then do;
        if ZLAB(i-1) ne .                    then ZLABB6_FV(i-45) = ZLABB6_FV(i-46);
        else if ZLAB(i-2) ne .                then ZLABB6_FV(i-45) = ZLABB6_FV(i-47);
        else if ZLAB(i-2) = .                  then ZLABB6_FV(i-45) = .;
    end;
end;
do i=61 to 75;
    if ZLAB(i) ne .                        then ZLABB7_FV(i-60) = ZLAB(i);
    else if i = 61                          then ZLABB7_FV(i-60) = ZLABB7_FV0;
    else if i = 62 and ZLAB(i-1) ne . then ZLABB7_FV(i-60) = ZLABB7_FV1;
    else if i > 62 then do;
        if ZLAB(i-1) ne .                    then ZLABB7_FV(i-60) = ZLABB7_FV(i-61);
        else if ZLAB(i-2) ne .                then ZLABB7_FV(i-60) = ZLABB7_FV(i-62);
        else if ZLAB(i-2) = .                  then ZLABB7_FV(i-60) = .;
    end;
end;
do i=76 to 90;
    if ZLAB(i) ne .                        then ZLABB8_FV(i-75) = ZLAB(i);
    else if i = 76                          then ZLABB8_FV(i-75) = ZLABB8_FV0;
    else if i = 77 and ZLAB(i-1) ne . then ZLABB8_FV(i-75) = ZLABB8_FV1;
    else if i > 77 then do;
        if ZLAB(i-1) ne .                    then ZLABB8_FV(i-75) = ZLABB8_FV(i-76);
        else if ZLAB(i-2) ne .                then ZLABB8_FV(i-75) = ZLABB8_FV(i-77);
        else if ZLAB(i-2) = .                  then ZLABB8_FV(i-75) = .;
    end;
end;
do i=91 to 105;
    if ZLAB(i) ne .                        then ZLABB9_FV(i-90) = ZLAB(i);
    else if i = 91                          then ZLABB9_FV(i-90) = ZLABB9_FV0;
    else if i = 92 and ZLAB(i-1) ne . then ZLABB9_FV(i-90) = ZLABB9_FV1;
    else if i > 92 then do;
        if ZLAB(i-1) ne .                    then ZLABB9_FV(i-90) = ZLABB9_FV(i-91);
        else if ZLAB(i-2) ne .                then ZLABB9_FV(i-90) = ZLABB9_FV(i-92);
        else if ZLAB(i-2) = .                  then ZLABB9_FV(i-90) = .;
    end;
end;
end;

```

```

do i=106 to 120;
    if ZLAB(i) ne . then ZLABB10_FV(i-105) = ZLAB(i);
    else if i = 106 then ZLABB10_FV(i-105) = ZLABB10_FV0;
    else if i = 107 and ZLAB(i-1) ne . then ZLABB10_FV(i-105) = ZLABB10_FV1;
    else if i > 107 then do;
        if ZLAB(i-1) ne . then ZLABB10_FV(i-105) = ZLABB10_FV(i-106);
        else if ZLAB(i-2) ne . then ZLABB10_FV(i-105) = ZLABB10_FV(i-107);
        else if ZLAB(i-2) = . then ZLABB10_FV(i-105) = .;
    end;
end;
end;

```

```

HiBili = .;
if ZLABB1 ge 0.5 then HiBili = 1;
if . < ZLABB1 < 0.5 then HiBili = 0;
array HiBili_FV(15) HiBili_FV1-HiBili_FV15;
HiBili_FV0 = HiBili;
do i=1 to 15;
    if ZLABB1_FV(i) ge 0.5 then HiBili_FV(i) = 1;
    if . < ZLABB1_FV(i) < 0.5 then HiBili_FV(i) = 0;
end;

```

```

HiSerChol = .;
if ZLABB4 ge 250 then HiSerChol = 1;
if . < ZLABB4 < 250 then HiSerChol = 0;
array HiSerChol_FV(15) HiSerChol_FV1-HiSerChol_FV15;
HiSerChol_FV0 = HiSerChol;
do i=1 to 15;
    if ZLABB4_FV(i) ge 250 then HiSerChol_FV(i) = 1;
    if . < ZLABB4_FV(i) < 250 then HiSerChol_FV(i) = 0;
end;

```

```

HiSerTrigly = .;
if ZLABB5 ge 5.0 then HiSerTrigly = 1;
if . < ZLABB5 < 5.0 then HiSerTrigly = 0;
array HiSerTrigly_FV(15) HiSerTrigly_FV1-HiSerTrigly_FV15;
HiSerTrigly_FV0 = HiSerTrigly;
do i=1 to 15;
    if ZLABB5_FV(i) ge 5.0 then HiSerTrigly_FV(i) = 1;
    if . < ZLABB5_FV(i) < 5.0 then HiSerTrigly_FV(i) = 0;
end;

```

```

HiSerUric = .;
if ZLABB6 ge 7.0 then HiSerUric = 1;
if . < ZLABB6 < 7.0 then HiSerUric = 0;
array HiSerUric_FV(15) HiSerUric_FV1-HiSerUric_FV15;
HiSerUric_FV0 = HiSerUric;
do i=1 to 15;

```

```

        if ZLABB6_FV(i) ge 7.0          then HiSerUric_FV(i) = 1;
        if . < ZLABB6_FV(i) < 7.0      then HiSerUric_FV(i) = 0;
    end;

```

```

HiSerAlk = .;
if ZLABB7 ge 7.5          then HiSerAlk = 1;
if . < ZLABB7 < 7.5      then HiSerAlk = 0;
array HiSerAlk_FV(15) HiSerAlk_FV1-HiSerAlk_FV15;
HiSerAlk_FV0 = HiSerAlk;
do i=1 to 15;
    if ZLABB7_FV(i) ge 7.5          then HiSerAlk_FV(i) = 1;
    if . < ZLABB7_FV(i) < 7.5      then HiSerAlk_FV(i) = 0;
end;

```

```

HiPlasUrea = .;
if ZLABB8 ge 16 then HiPlasUrea = 1;
if . < ZLABB8 < 16      then HiPlasUrea = 0;
array HiPlasUrea_FV(15) HiPlasUrea_FV1-HiPlasUrea_FV15;
HiPlasUrea_FV0 = HiPlasUrea;
do i=1 to 15;
    if ZLABB8_FV(i) ge 16          then HiPlasUrea_FV(i) = 1;
    if . < ZLABB8_FV(i) < 16 then HiPlasUrea_FV(i) = 0;
end;

```

```

HiFastGluc = .;
if ZLABB9 ge 100          then HiFastGluc = 1;
if . < ZLABB9 < 100      then HiFastGluc = 0;
array HiFastGluc_FV(15) HiFastGluc_FV1-HiFastGluc_FV15;
HiFastGluc_FV0 = HiFastGluc;
do i=1 to 15;
    if ZLABB9_FV(i) ge 100          then HiFastGluc_FV(i) = 1;
    if . < ZLABB9_FV(i) < 100      then HiFastGluc_FV(i) = 0;
end;

```

```

HiOneGluc = .;
if ZLABB10 ge 180          then HiOneGluc = 1;
if . < ZLABB10 < 180      then HiOneGluc = 0;
array HiOneGluc_FV(15) HiOneGluc_FV1-HiOneGluc_FV15;
HiOneGluc_FV0 = HiOneGluc;
do i=1 to 15;
    if ZLABB10_FV(i) ge 180          then HiOneGluc_FV(i) = 1;
    if . < ZLABB10_FV(i) < 180      then HiOneGluc_FV(i) = 0;
end;

```

```

keep    HiBili          HiBili_FV0-HiBili_Fv15
        HiSerChol       HiSerChol_FV0-HiSerChol_FV15
        HiSerTrigly     HiSerTrigly_FV0-HiSerTrigly_FV15
        HiSerUric       HiSerUric_FV0-HiSerUric_FV15

```

```

        HiSerAlk      HiSerAlk_FV0-HiSerAlk_FV15
        HiPlasUrea    HiPlasUrea_FV0-HiPlasUrea_FV15
        HiFastGluc    HiFastGluc_FV0-HiFastGluc_FV15
        HiOneGluc     HiOneGluc_FV0-HiOneGluc_FV15
    ;

drop ZLAB4-ZLAB18

                                ZLAB91-ZLAB105
                                ZLAB120-ZLAB134
                                ZLAB149-ZLAB163
                                ZLAB178-ZLAB192
                                ZLAB207-ZLAB221
                                ZLAB236-ZLAB250
                                ZLAB265-ZLAB279;

array WBC1_FV(17) WBC1_FV1 - WBC1_FV17;
array WBC12_FV(17) WBC12_FV1 - WBC12_FV17;
array WBC23_FV(17) WBC23_FV1 - WBC23_FV17;
array WBC(33) WBC1-WBC33;
WBC1_FV0 = WBC1;
WBC1_FV1 = WBC1;
WBC1_FV2 = WBC1;
do i = 2 to 6;
    if WBC(i) ne . then do;
        WBC1_FV((i-1)*3) = WBC(i);
        WBC1_FV((i-1)*3+1) = WBC(i);
        WBC1_FV((i-1)*3+2) = WBC(i);
    end;
    else do;
        WBC1_FV((i-1)*3) = .;
        WBC1_FV((i-1)*3+1) = .;
        WBC1_FV((i-1)*3+2) = .;
    end;
end;
WBC12_FV0 = WBC12;
WBC12_FV1 = WBC12;
WBC12_FV2 = WBC12;
do i = 13 to 17;
    if WBC(i) ne . then do;
        WBC12_FV((i-12)*3) = WBC(i);
        WBC12_FV((i-12)*3+1) = WBC(i);
        WBC12_FV((i-12)*3+2) = WBC(i);
    end;
    else do;
        WBC12_FV((i-12)*3) = .;
        WBC12_FV((i-12)*3+1) = .;
        WBC12_FV((i-12)*3+2) = .;
    end;
end;

```

```

end;
WBC23_FV0 = WBC23;
WBC23_FV1 = WBC23;
WBC23_FV2 = WBC23;
do i = 24 to 28;
    if WBC(i) ne . then do;
        WBC23_FV((i-23)*3) = WBC(i);
        WBC23_FV((i-23)*3+1) = WBC(i);
        WBC23_FV((i-23)*3+2) = WBC(i);
    end;
    else do;
        WBC23_FV((i-23)*3) = .;
        WBC23_FV((i-23)*3+1) = .;
        WBC23_FV((i-23)*3+2) = .;
    end;
end;
end;

```

```

HiWhiteCell = .;
if WBC1 ge 7500 then HiWhiteCell = 1;
if . < WBC1 < 7500 then HiWhiteCell = 0;
array HiWhiteCell_FV(15) HiWhiteCell_FV1-HiWhiteCell_FV15;
HiWhiteCell_FV0 = HiWhiteCell;
do i=1 to 15;
    if WBC1_FV(i) ge 7500 then HiWhiteCell_FV(i) = 1;
    if . < WBC1_FV(i) < 7500 then HiWhiteCell_FV(i) = 0;
end;

```

```

HiNeut = .;
if WBC12 ge 4500 then HiNeut = 1;
if . < WBC12 < 4500 then HiNeut = 0;
array HiNeut_FV(15) HiNeut_FV1-HiNeut_FV15;
HiNeut_FV0 = HiNeut;
do i=1 to 15;
    if WBC12_FV(i) ge 4500 then HiNeut_FV(i) = 1;
    if . < WBC12_FV(i) < 4500 then HiNeut_FV(i) = 0;
end;

```

```

HiHemat = .;
if WBC23 ge 46 then HiHemat = 1;
if . < WBC23 < 46 then HiHemat = 0;
array HiHemat_FV(15) HiHemat_FV1-HiHemat_FV15;
HiHemat_FV0 = HiHemat;
do i=1 to 15;
    if WBC23_FV(i) ge 46 then HiHemat_FV(i) = 1;
    if . < WBC23_FV(i) < 46 then HiHemat_FV(i) = 0;
end;

```



```

keep HiWhiteCell HiWhiteCell_FV0-HiWhiteCell_FV15
      HiNeut      HiNeut_FV0-HiNeut_FV15
      HiHemat     HiHemat_FV0-HiHemat_FV15
      ;

```

```

drop WBC1-WBC33;

```

```

/*Other covariates*/

```

```

array ICG_FV(17) ICG_FV1 - ICG_FV17;
array ICG(6) ICG1 - ICG6;
ICG_FV0 = ICG1;
ICG_FV1 = ICG1;
ICG_FV2 = ICG1;
do i = 2 to 6;
    if ICG(i) ne . then do;
        ICG_FV((i-1)*3) = ICG(i);
        ICG_FV((i-1)*3+1) = ICG(i);
        ICG_FV((i-1)*3+2) = ICG(i);
    end;
    else do;
        ICG_FV((i-1)*3) = .;
        ICG_FV((i-1)*3+1) = .;
        ICG_FV((i-1)*3+2) = .;
    end;
end;

```

```

end;

```

```

CIG = .;
if ICG1 = 1 then CIG = 0;
if ICG1 in (2,3,4,5,6) then CIG = 1;
keep CIG;
array CIG_FV(15) CIG_FV1 - CIG_FV15;
CIG_FV0 = CIG;
do i = 1 to 15;
    if ICG_FV(i) = 1 then CIG_FV(i) = 0;
    if ICG_FV(i) in (2,3,4,5,6) then CIG_FV(i) = 1;
end;
keep CIG_FV0 - CIG_FV15;
drop ICG1 - ICG6;

```

```

array IJOB_FV(17) IJOB_FV1 - IJOB_FV17;
array IJOB(6) IJOB34 - IJOB39;
IJOB_FV0 = IJOB34;
IJOB_FV1 = IJOB34;
IJOB_FV2 = IJOB34;
do i = 2 to 6;

```

```

        if IJOB(i) ne . then do;
            IJOB_FV((i-1)*3) = IJOB(i);
            IJOB_FV((i-1)*3+1) = IJOB(i);
            IJOB_FV((i-1)*3+2) = IJOB(i);
        end;
    else do;
        IJOB_FV((i-1)*3) = .;
        IJOB_FV((i-1)*3+1) = .;
        IJOB_FV((i-1)*3+2) = .;
    end;
end;

```

```

INACT = .;
if IJOB34 = 1 then INACT = 1;
if IJOB34 = 2 then INACT = 0;
if IJOB34 = 3 then INACT = 0;
keep INACT;
array INACT_FV(15) INACT_FV1 - INACT_FV15;
INACT_FV0 = INACT;
do i=1 to 15;
    if IJOB_FV(i) = 1 then INACT_FV(i) = 1;
    if IJOB_FV(i) = 2 then INACT_FV(i) = 0;
    if IJOB_FV(i) = 3 then INACT_FV(i) = 0;
end;
keep INACT_FV0-INACT_FV15;
drop IJOB34 - IJOB39;

```

```

/*****Adherence measures*****/

```

```

/*Notes on visit number: */
/*icap1-26 prescription made at FV1-26*/
/*caps1-26 prescription made for baseline, FV1-25*/
/*iadh1-26; adh1-26 adherence observed at baseline, fv1 - 25*/
array inv_a(*) inv1 - inv24;
array iadh(26) iadh1-iadh26;
array caps(26) caps1-caps26;
array capsFV(*) capsFV1 - capsFV26;
array adh(52) adh1-adh52;
array adhFV(26) adhFV0 - adhFV25;

```

```

/*clear adherence variable to recalculate*/
do i = 2 to 52;
    adh(i) = .;
end;

```

```

/*Pre-baseline adherence: adherence to placebo assessed pre-randomization for all participants*/

```

```

if JADH1 ne . and JADH2 ne . then adhpre0 = (JADH1 + JADH2)/2;
else if JADH1 = . and JADH2 ne . then adhpre0 = JADH2;
else if JADH1 ne . and JADH2 = . then adhpre0 = JADH1;
else if JADH1 = . and JADH2 = . then adhpre0 = .;

```

```

adhpre0bin = .;
if adhpre0 ge 80 then adhpre0bin = 0;
if .< adhpre0 <80 then adhpre0bin = 1;

```

```
/*Follow-up: calculate adherence at each time point*/
```

```

do i = 1 to 26;
  capsFV(i) = caps(i);
  if caps(i) = . then do;
    if i = 2 then capsFV(i) = 9;
    else if caps(i-1) ne . then capsFV(i) = caps(i-1);
    else if caps(i-1) = . then do;
      if i > 2 then capsFV(i) = caps(i-2);
    end;
  end;
end;

```

```

do i=2 to 26;
  if invdth = 0 or i le invdth then do;
    if . < iadh(i) < 6 then adh(i)=(capsFV(i)/9)*(110-iadh(i)*20);
    else if iadh(i)= 6 then adh(i)=0; /*if no prescription, then adh = 0%*/
  end;
  else if i > invdth then adh(i) = .;
end;

```

```

/*carry forward adherence for a maximum of three consecutive missed visits (carry forward 3 times
because occurs b/f visit when measured);*/
array indic(*) indic1-indic15;
indic0 = 0;

```

```

do i = 1 to 15;
  if adh(i) ne . then adhFV(i) = adh(i);
  else if i = 1 then adhFV(i) = .;
  else if i = 2 then adhFV(i) = adhFV(i-1);
  else if i = 3 then do;
    if adh(i-1) ne . then adhFV(i) = adh(i-1);
    else if adh(i-2) ne . then adhFV(i) = adh(i-2);
    else adhFV(i) = .;
  end;
  else if i = 4 then do;
    if adh(i-1) ne . then adhFV(i) = adh(i-1);
    else if adh(i-2) ne . then adhFV(i) = adh(i-2);
  end;

```

```

                else if adh(i-2) = .      then adhFV(i) = .;    /*special case of censoring at visit 4
because Adh0 can be measured at IV4 or IV5*/
            end;
            else if i > 4      then do;
                if adh(i-1) ne . then adhFV(i) = adh(i-1);
                else if adh(i-2) ne . then adhFV(i) = adh(i-2);
                else if adh(i-3) ne . then adhFV(i) = adh(i-3);
                else if adh(i-3) = .      then adhFV(i) = .;
            end;

            if i in (1,2,3)                                then indic(i) = 0;
            else if adh(i) = . and adh(i-1) = . and adh(i-2) ne . then indic(i) = 1;
            else if adh(i) = . and adh(i-1) = . and adh(i-2) = . then indic(i) = .;
            else                                             indic(i) = 0;

            if invdth ne 0 and i ge invdth then do;
                indic(i) = .;
                adhFV(i) = .;
            end;
        end;

keep adhFV0 -adhFV25 indic0-indic15 ;

/*censoring indicator - adherence*/
/*censvisit = visit # when adherence first is missing or end of 5-year follow-up*/
censvisit = .;
do i = 1 to 15; /*adh15 is adherence measured at visit 15 for the intervisit period 14-15 (last adherence
before 5 years)*/
    if censvisit ne . then censvisit = censvisit;
    else if censvisit = . then do;
        if adhFV(i) = . then censvisit = i-1;
        else if i < 15 and adhFV(i) ne . then censvisit = .;
        else if i = 15 and adhFV(i) ne . then censvisit = 15;
    end;
end;
keep censvisit;

/*For i = 27-52: ADH(i) = cumulative adherence from baseline to FV(x), x = 1-15*/
adh27=adhFV0;

do i=2 to 15;
    if adhFV(i) ne . then adh(26+i)=(adh(26+i-1)*(i-1)+adhFV(i))/i;
    else adh(26+i) = .;
end;

array adhbin_array(*)          adhbin0-adhbin25 adhbin27-adhbin42; /*binary time-varying
adherence level*/

```

```

if . < adhFV0 < 80 then adhbin0 = 1;
else if adhFV0 >=80 then adhbin0 = 0;
else if adhFV0 = . then adhbin0 = .;

do i=2 to 15;
    if . < adh(i) < 80 then adhbin_array(i) = 1;
    else if adh(i) >=80 then adhbin_array(i) = 0;
    else if adh(i) = . then adhbin_array(i) = .;
end;
adhbin27 = adhbin0;
do i=28 to 42;
    if . < adh(i) < 80 then adhbin_array(i) = 1;
    else if adh(i) >=80 then adhbin_array(i) = 0;
    else if adh(i) = . then adhbin_array(i) = .;
end;
keep adhbin0-adhbin15 adhbin27-adhbin42;

if dth5 = 1 and invdth = 0 then invdth = ceil(min(itd,itdx)*3/365);

/*New variable: ADHX15. Cumulative average adherence at last visit before death or 5 years, whichever
is first*/
if dth5 = 1 and invdth = 1 then adh15 = adh27;
else if dth5 = 1 and invdth > 0 then adh15 = adh(26+(invdth-1));
else if dth5 = 1 and invdth = 0 and 0 < inv28 le 15 then adh15 = adh(26+(inv_a(inv28)));
else if dth5 ne 1 then adh15 = adh(26+15);

Adhx15Bin = .;
if Adhx15 ge 80 then Adhx15Bin = 0;
if .<Adhx15 <80 then Adhx15Bin = 1;
keep Adhx15Bin adhx15;

keep adhpre0bin adh1-adh52 ;

/****Code adherence to match original paper, where adherence = 0 on all missed visits*****/
array inv(*) inv1-inv27;
array old_adh(52) old_adh1-old_adh52;

old_adh1 = adh1;
if adh1 = . then old_adh1 = 0; /*set adherence to 0 when missing*/

ihigh=2;
if invdth=0 and inv28>14 then ihigh=inv28;
if (istat=3 or istat=6) and lvss>ihigh then ihigh=lvss;
if invdth>2 then ihigh=invdth-1;

old_adh27=old_adh1;

```

```

do i=2 to ihigh;
    if iadh(i)=6 then old_adh(i)=0; /*if no prescription, then adh =0%*/
    else if inv(i)=0 or inv(i)=2 or inv(i)=7 or inv(i)=8 then old_adh(i)=0;
    else if capsFV(i) = . or iadh(i) = . then old_adh(i) = 0;
    else old_adh(i)= adh(i);
    if old_adh(i) = . then old_adh(i) = 0;
end;
do i=2 to ihigh;
    old_adh(26+i)=(old_adh(26+i-1)*(i-1)+old_adh(i))/i;
end;
keep old_adh1-old_adh52;

old_adhx=old_adh(26+ihigh);
old_adhx15=old_adhx;
if ihigh>15 then old_adhx15=old_adh41;

old_Adhx15Bin = .;
if old_Adhx15 ge 80 then old_Adhx15Bin = 0;
if .<old_Adhx15 <80 then old_Adhx15Bin = 1;
keep old_Adhx15Bin old_adhx old_adhx15;

run;

proc freq data = cdp.binary nlevels;
tables ID /noprint;
title 'Before exclusion of missing baseline vars';
run;

data cdp.binary;
set cdp.binary;
where adhbino ne . and age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
    NIHA_FV0 ne . and HiSysBP_FV0 ne . and HiDiasBP_FV0 ne . and
    HiWhiteCell_FV0 ne . and HiNeut_FV0 ne . and HiHemat_FV0 ne . and
    HiBili_FV0 ne . and HiSerChol_FV0 ne . and HiSerTrigly_FV0 ne . and
    HiSerUric_FV0 ne . and HiSerAlk_FV0 ne . and HiPlasUrea_FV0 ne . and
    HiFastGluc_FV0 ne . and HiOneGluc_FV0 ne . and HiHeart_FV0 ne . and
    CHF_FV0 ne . and ACI_FV0 ne . and AP_FV0 ne . and
    IC_FV0 ne . and ICIA_FV0 ne . and DIG_FV0 ne . and DIUR_FV0 ne . and
    AntiArr_FV0 ne . and AntiHyp_FV0 ne . and OralHyp_FV0 ne . and
    CardioM_FV0 ne . and AnyQQS_FV0 ne . and AnySTDep_FV0 ne . and
    AnyTWave_FV0 ne . and
    STElev_FV0 ne . and FVEB_FV0 ne . and VCD_FV0 ne . and CIG_FV0 ne . and
    INACT_FV0 ne .;
run;

proc freq data = cdp.binary nlevels;
tables ID /noprint;

```

```
title 'After exclusion of missing baseline vars';
run;
```

```
/******  
**Long format dataset for analysis of*****  
**Censoring-Weighted Cumulative Adherence*****  
*****
```

```
/*Convert data into observations for each follow-up visit*/  
/*For censoring due to change in missing adherence only*/  
/*As with baseline-only analysis, individuals with missing*/  
/*covariates over time not censored - covariates carried */  
/*forward from last recorded value for censoring weights*/
```

```
data cdp.censwt_ag;  
set cdp.binary;
```

```
array adhbin_a(*)          adhbin0 - adhbin15;  
array adh(*)               adh1-adh16;  
array indic(*)             indic0 - indic15;
```

```
array NIHA_FV_a(*)         NIHA_FV0 - NIHA_FV15;  
array HiSysBP_FV_a(*)      HiSysBP_FV0-HiSysBP_FV15;  
array HiDiasBP_FV_a(*)     HiDiasBP_FV0-HiDiasBP_FV15 ;  
array HiWhiteCell_FV_a(*)  HiWhiteCell_FV0-HiWhiteCell_FV15;  
array HiNeut_FV_a(*)       HiNeut_FV0-HiNeut_FV15;  
array HiHemat_FV_a(*)      HiHemat_FV0-HiHemat_FV15;  
array HiBili_FV_a(*)       HiBili_FV0-HiBili_FV15;  
array HiSerChol_FV_a(*)    HiSerChol_FV0-HiSerChol_FV15;  
array HiSerTrigly_FV_a(*)  HiSerTrigly_FV0-HiSerTrigly_FV15;  
array HiSerUric_FV_a(*)    HiSerUric_FV0-HiSerUric_FV15;  
array HiSerAlk_FV_a(*)     HiSerAlk_FV0-HiSerAlk_FV15;  
array HiPlasUrea_FV_a(*)   HiPlasUrea_FV0-HiPlasUrea_FV15;  
array HiFastGluc_FV_a(*)   HiFastGluc_FV0-HiFastGluc_FV15;  
array HiOneGluc_FV_a(*)    HiOneGluc_FV0-HiOneGluc_FV15;  
array HiHeart_FV_a(*)      HiHeart_FV0-HiHeart_FV15;  
array CHF_FV_a(*)          CHF_FV0-CHF_FV15;  
array ACI_FV_a(*)          ACI_FV0-ACI_FV15;  
array AP_FV_a(*)           AP_FV0-AP_FV15;  
array IC_FV_a(*)           IC_FV0-IC_FV15;  
array ICIA_FV_a(*)         ICIA_FV0-ICIA_FV15;  
array DIG_FV_a(*)          DIG_FV0-DIG_FV15;  
array DIUR_FV_a(*)         DIUR_FV0-DIUR_FV15;  
array AntiArr_FV_a(*)      AntiArr_FV0-AntiArr_FV15;  
array AntiHyp_FV_a(*)      AntiHyp_FV0-AntiHyp_FV15;  
array OralHyp_FV_a(*)      OralHyp_FV0-OralHyp_FV15;  
array CardioM_FV_a(*)      CardioM_FV0 - CardioM_FV15;  
array AnyQQS_FV_a(*)       AnyQQS_FV0-AnyQQS_FV15;
```

```

array AnySTDep_FV_a(*)          AnySTDep_FV0-AnySTDep_FV15;
array AnyTwave_FV_a(*)          AnyTwave_FV0-AnyTwave_FV15;
array STElev_FV_a(*)            STElev_FV0-STElev_FV15;
array FVEB_FV_a(*)              FVEB_FV0-FVEB_FV15;
array VCD_FV_a(*)               VCD_FV0-VCD_FV15;
array CIG_FV_a(*)               CIG_FV0 - CIG_FV15;
array INACT_FV_a(*)             INACT_FV0-INACT_FV15;

```

```
do i = 1 to 15;
```

```
    ind = indic(i);
```

```
    visit = i-1; /*visit number: 0 = baseline, 1 = FV1, etc*/
```

```
    if i < 4 then do;
```

```
        if invdth = 0 then do;
```

```
            cens = -1;
```

```
            p_cens0 = 1;
```

```
            adhr_t = adhbin_a(i);
```

```
            death = 0;
```

```
            if i ge 2 then adhr_t1 = adhbin_a(i-1);
```

```
            if i ge 3 then adhr_t2 = adhbin_a(i-2);
```

```
        end;
```

```
    else if i > invdth then do;
```

```
        cens = .;
```

```
        p_cens0 = .;
```

```
        adhr_t = .;
```

```
        death = .;
```

```
    end;
```

```
    else if i le invdth then do;
```

```
        cens = -1;
```

```
        p_cens0 = 1;
```

```
        adhr_t = adhbin_a(i);
```

```
        if invdth = i then death = 1;
```

```
        else if invdth > i then death = 0;
```

```
        if death in (0,1) then do;
```

```
            if i ge 2 then adhr_t1 = adhbin_a(i-1);
```

```
            if i ge 3 then adhr_t2 = adhbin_a(i-2);
```

```
        end;
```

```
    end;
```



```

end;

else if i ge 4 then do;
    if indic(i) = 0 then do;
        cens = -1;
        p_cens0 = 1;

        if invdth = i then death = 1;
        else if invdth = 0 or invdth > i then death = 0;
    end;
    else if indic(i) = 1 then do;
        if adh(i) = . then cens = 1;
        else if adh(i) ne . then cens = 0;
        p_cens0 = .;
    end;
    else if indic(i) = . then do;
        cens = .;
        p_cens0 = .;
        death = .;
    end;

    /*current and baseline adherence*/
    if death in (0,1) then do;
        adhr_t = adhbin_a(i);

        adhr_t1 = adhbin_a(i-1);
        adhr_t2 = adhbin_a(i-2);
        adhr_t3 = adhbin_a(i-3);
    end;
    else if death = . then do;
        adhr_t = .;

        adhr_t1 = .;
        adhr_t2 = .;
        adhr_t3 = .;
    end;
end;

/*if death = 1 and i > 1 then adhr_t = adhbin_a(i-1);*/

if death in (0,1) then do;
    if adhr_t = . then adh_measure = 0;
    else if adhr_t ne . then adh_measure = 1;
end;
else if death = . then adh_measure = .;

/*Save covariates for visit i*/
NIHA_bin0 = NIHA_FV0;

```

```

NIHAFV = NIHA_FV_a(i);
if i > 1 then NIHAFV_t1 = NIHA_FV_a(i-1);      else NIHAFV_t1 = NIHA_bin0;
    if NIHAFV_t1 = . then NIHAFV_t1 = NIHA_FV_a(i-2);
    if NIHAFV_t1 = . then NIHAFV_t1 = NIHA_FV_a(i-3);

```

```

CHF0 = CHF_FV0;
ACI0 = ACI_FV0;
AP0 = AP_FV0;
IC0 = IC_FV0;
ICIA0 = ICIA_FV0;

```

```

CHFFV = CHF_FV_a(i);
ACIFV = ACI_FV_a(i);
APFV = AP_FV_a(i);
ICFV = IC_FV_a(i);
ICIAFV = ICIA_FV_a(i);

```

```

if i > 1 then CHFFV_t1 = CHF_FV_a(i-1); else CHFFV_t1 = CHF0;
if i > 1 then ACIFV_t1 = ACI_FV_a(i-1);      else ACIFV_t1 = ACI0;
if i > 1 then APFV_t1 = AP_FV_a(i-1);      else APFV_t1 = AP0;
if i > 1 then ICFV_t1 = IC_FV_a(i-1);      else ICFV_t1 = IC0;
if i > 1 then ICIAFV_t1 = ICIA_FV_a(i-1); else ICIAFV_t1 = ICIA0;

```

```

    if CHFFV_t1 = . then CHFFV_t1 = CHF_FV_a(i-2);
    if ACIFV_t1 = . then ACIFV_t1 = ACI_FV_a(i-2);
    if APFV_t1 = . then APFV_t1 = AP_FV_a(i-2);
    if ICFV_t1 = . then ICFV_t1 = IC_FV_a(i-2);
    if ICIAFV_t1 = . then ICIAFV_t1 = ICIA_FV_a(i-2);

```

```

DIG0 = DIG_FV0;
DIUR0 = DIUR_FV0;
AntiArr0 = AntiARR_FV0;
AntiHyp0 = AntiHyp_FV0;
OralHyp0 = OralHyp_FV0;

```

```

DIGFV = DIG_FV_a(i);
DIURFV = DIUR_FV_a(i);
AntiArrFV = AntiARR_FV_a(i);
AntiHypFV = AntiHyp_FV_a(i);
OralHypFV = OralHyp_FV_a(i);

```

```

if i > 1 then DIGFV_t1 = DIG_FV_a(i-1);      else DigFV_t1 = DIG0;
if i > 1 then DIURFV_t1 = DIUR_FV_a(i-1);    else DiurFV_t1 = DIUR0;
if i > 1 then AntiArrFV_t1 = AntiARR_FV_a(i-1); else AntiArrFV_t1 = AntiArr0;
if i > 1 then AntiHypFV_t1 = AntiHyp_FV_a(i-1); else AntiHypFV_t1 = AntiHyp0;
if i > 1 then OralHypFV_t1 = OralHyp_FV_a(i-1); else OralHypFV_t1 = OralHyp0;

```

```

if DIGFV_t1 = . then DIGFV_t1 = DIG_FV_a(i-2);
if DIURFV_t1 = . then DIURFV_t1 = DIUR_FV_a(i-2);
if AntiArrFV_t1 = . then AntiArrFV_t1 = AntiARR_FV_a(i-2);
if AntiHypFV_t1 = . then AntiHypFV_t1 = AntiHyp_FV_a(i-2);
if OralHypFV_t1 = . then OralHypFV_t1 = OralHyp_FV_a(i-2);

```

```

CardioM0 = CardioM_FV0;
CardioMFV = CardioM_FV_a(i);
if i > 1 then CardioMFV_t1 = CardioM_FV_a(i-1);      else CardioMFV_t1 = CardioM0;
    if CardioMFV_t1 = . then CardioMFV_t1 = CardioM_FV_a(i-2);

```

```

AnyQQS0 = AnyQQS_FV0;
AnySTDep0 = AnySTDep_FV0;
AnyTWave0 = AnyTWave_FV0;
STelev0 = STElev_FV0;
FVEB0 = FVEB_FV0;
VCD0 = VCD_FV0;
HiHeart0 = HiHeart_FV0;

```

```

AnyQQSFV = AnyQQS_FV_a(i);
AnySTDepFV = AnySTDep_FV_a(i);
AnyTWaveFV = AnyTWave_FV_a(i);
STelevFV = STElev_FV_a(i);
FVEBFV = FVEB_FV_a(i);
VCDFV = VCD_FV_a(i);
HiHeartFV = HiHeart_FV_a(i);

```

```

if i > 1 then AnyQQSFV_t1 = AnyQQS_FV_a(i-1);      else AnyQQSFV_t1 = AnyQQS0;
if i > 1 then AnySTDepFV_t1 = AnySTDep_FV_a(i-1);  else AnyStDepFV_t1 = AnyStDep0;
if i > 1 then AnyTWaveFV_t1 = AnyTWave_FV_a(i-1);  else AnyTWaveFV_t1 = AnyTWave0;
if i > 1 then STElevFV_t1 = STElev_FV_a(i-1);      else StElevFV_t1 = StElev0;
if i > 1 then FVEBFV_t1 = FVEB_FV_a(i-1);          else FVEBFV_t1 = FVEB0;
if i > 1 then VCDFV_t1 = VCD_FV_a(i-1);            else VCDFV_t1 = VCD0;
if i > 1 then HiHeartFV_t1 = HiHeart_FV_a(i-1);    else HiHeartFV_t1 = HiHeart0;

```

```

if AnyQQSFV_t1 = . then AnyQQSFV_t1 = AnyQQS_FV_a(i-2);
if AnySTDepFV_t1 = . then AnySTDepFV_t1 = AnySTDep_FV_a(i-2);
if AnyTWaveFV_t1 = . then AnyTWaveFV_t1 = AnyTWave_FV_a(i-2);
if STElevFV_t1 = . then STElevFV_t1 = STElev_FV_a(i-2);
if FVEBFV_t1 = . then FVEBFV_t1 = FVEB_FV_a(i-2);
if VCDFV_t1 = . then VCDFV_t1 = VCD_FV_a(i-2);
if HiHeartFV_t1 = . then HiHeartFV_t1 = HiHeart_FV_a(i-2);

```

```

HiBili0 = HiBili_FV0;
HiSerChol0 = HiSerChol_FV0;
HiSerTrigly0 = HiSerTrigly_FV0;
HiSerUric0 = HiSerUric_FV0;
HiSerAlk0 = HiSerAlk_FV0;
HiPlasUrea0 = HiPlasUrea_FV0;
HiFastGluc0 = HiFastGluc_FV0;
HiOneGluc0 = HiOneGluc_FV0;

```

```

HiBiliFV = HiBili_FV_a(i);
HiSerCholFV = HiSerChol_FV_a(i);
HiSerTriglyFV = HiSerTrigly_FV_a(i);
HiSerUricFV = HiSerUric_FV_a(i);
HiSerAlkFV = HiSerAlk_FV_a(i);
HiPlasUreaFV = HiPlasUrea_FV_a(i);
HiFastGlucFV = HiFastGluc_FV_a(i);
HiOneGlucFV = HiOneGluc_FV_a(i);

```

```

if i > 1 then HiBiliFV_t1 = HiBili_FV_a(i-1);           else HiBiliFV_t1 = HiBili0;
if i > 1 then HiSerCholFV_t1 = HiSerChol_FV_a(i-1);    else HiSerCholFV_t1 = HiSerChol0;
if i > 1 then HiSerTriglyFV_t1 = HiSerTrigly_FV_a(i-1); else HiSerTriglyFV_t1 = HiSerTrigly0;
if i > 1 then HiSerUricFV_t1 = HiSerUric_FV_a(i-1);    else HiSerUricFV_t1 = HiSerUric0;
if i > 1 then HiSerAlkFV_t1 = HiSerAlk_FV_a(i-1); else HiSerAlkFV_t1 = HiSerAlk0;
if i > 1 then HiPlasUreaFV_t1 = HiPlasUrea_FV_a(i-1);  else HiPlasUreaFV_t1 = HiPlasUrea0;
if i > 1 then HiFastGlucFV_t1 = HiFastGluc_FV_a(i-1);  else HiFastGlucFV_t1 = HiFastGluc0;
if i > 1 then HiOneGlucFV_t1 = HiOneGluc_FV_a(i-1);    else HiOneGlucFV_t1 = HiOneGluc0;

```

```

if HiBiliFV_t1 = . then HiBiliFV_t1 = HiBili_FV_a(i-2);
if HiSerCholFV_t1 = . then HiSerCholFV_t1 = HiSerChol_FV_a(i-2);
if HiSerTriglyFV_t1 = . then HiSerTriglyFV_t1 = HiSerTrigly_FV_a(i-2);
if HiSerUricFV_t1 = . then HiSerUricFV_t1 = HiSerUric_FV_a(i-2);
if HiSerAlkFV_t1 = . then HiSerAlkFV_t1 = HiSerAlk_FV_a(i-2);
if HiPlasUreaFV_t1 = . then HiPlasUreaFV_t1 = HiPlasUrea_FV_a(i-2);
if HiFastGlucFV_t1 = . then HiFastGlucFV_t1 = HiFastGluc_FV_a(i-2);
if HiOneGlucFV_t1 = . then HiOneGlucFV_t1 = HiOneGluc_FV_a(i-2);

```

```

HiSysBP0 = HiSysBP_FV0;
HiDiasBP0 = HiDiasBP_FV0;

```

```

HiSysBPFV = HiSysBP_FV_a(i);
HiDiasBPFV = HiDiasBP_FV_a(i);

```

```

if i > 1 then HiSysBPFV_t1 = HiSysBP_FV_a(i-1);           else HiSysBPFV_t1 = HiSysBP0;
if i > 1 then HiDiasBPFV_t1 = HiDiasBP_FV_a(i-1);        else HiDiasBPFV_t1 = HiDiasBP0;

```

```

if HiSysBPFV_t1 = . then HiSysBPFV_t1 = HiSysBP_FV_a(i-2);
if HiDiasBPFV_t1 = . then HiDiasBPFV_t1 = HiDiasBP_FV_a(i-2);

```

```

CIG0 = CIG_FV0;
INACT0 = INACT_FV0;

```

```

CIGFV = CIG_FV_a(i);
INACTFV = INACT_FV_a(i);

```

```

if i > 1 then CIGFV_t1 = CIG_FV_a(i-1); else CIGFV_t1 = CIG0;
if i > 1 then INACTFV_t1 = INACT_FV_a(i-1); else InactFV_t1 = Inact0;

```

```

if CIGFV_t1 = . then CIGFV_t1 = CIG_FV_a(i-2);
if INACTFV_t1 = . then INACTFV_t1 = INACT_FV_a(i-2);

```

```

HiWhiteCell0 = HiWhiteCell_FV0;
HiNeut0      = HiNeut_FV0;
HiHemat0     = HiHemat_FV0;

```

```

HiWhiteCellFV = HiWhiteCell_FV_a(i);
HiNeutFV      = HiNeut_FV_a(i);
HiHematFV     = HiHemat_FV_a(i);

```

```

if i > 1 then HiWhiteCellFV_t1 = HiWhiteCell_FV_a(i-1); else HiWhiteCellFV_t1 = HiWhiteCell0;
if i > 1 then HiNeutFV_t1 = HiNeut_FV_a(i-1); else HiNeutFV_t1 = HiNeut0;
if i > 1 then HiHematFV_t1 = HiHemat_FV_a(i-1); else HiHematFV_t1 =
HiHemat0;

```

```

if HiWhiteCellFV_t1 = . then HiWhiteCellFV_t1 = HiWhiteCell_FV_a(i-2);
if HiNeutFV_t1 = . then HiNeutFV_t1 = HiNeut_FV_a(i-2);
if HiHematFV_t1 = . then HiHematFV_t1 = HiHemat_FV_a(i-2);

```

```

keep ID istat visit dth5 invdth
Adhx15Bin itr entrydate death

```

```

adhbino adhr_t adhr_t1 adhr_t2 adhr_t3 cens p_cens0 adh_measure

```

```

adh1-adh16 adhbino-adhbino15 indic0 - indic15 ind

```

```

adhpreno bin age_bin nonwhite IRK MI_bin RBW_bin age_cat

```

```
NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
IC0 ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STelev0 FVEB0 VCD0 CIG0 INACT0

NIHAFV HiSysBPFV HiDiasBPFV HiWhiteCellFV HiNeutFV HiHematFV
HiBiliFV HiSerCholFV HiSerTriglyFV HiSerUricFV HiSerAlkFV
HiPlasUreaFV HiFastGlucFV HiOneGlucFV HiHeartFV
CHFFV ACIFV APFV ICFV ICIAFV DIGFV DIURFV AntiArrFV
AntiHypFV OralHypFV CardioMFV AnyQQSFV AnySTDepFV
AnyTWaveFV STElevFV FVEBFV VCDFV
CIGFV INACTFV

NIHAFV_t1 HiSysBPFV_t1 HiDiasBPFV_t1 HiWhiteCellFV_t1 HiNeutFV_t1
HiHematFV_t1

HiBiliFV_t1 HiSerCholFV_t1 HiSerTriglyFV_t1 HiSerUricFV_t1 HiSerAlkFV_t1
HiPlasUreaFV_t1 HiFastGlucFV_t1 HiOneGlucFV_t1 HiHeartFV_t1
CHFFV_t1 ACIFV_t1 APFV_t1 ICFV_t1 ICIAFV_t1 DIGFV_t1 DIURFV_t1
AntiArrFV_t1

AntiHypFV_t1 OralHypFV_t1 CardioMFV_t1 AnyQQSFV_t1 AnySTDepFV_t1
AnyTWaveFV_t1 STElevFV_t1 FVEBFV_t1 VCDFV_t1
CIGFV_t1 INACTFV_t1 ;

output;
end;
run;

data cdp.censwt_ag;
set cdp.censwt_ag;

if death ne . then output;
run;

/*****/
/**Program 2: Replicate Analysis of 1980 NEJM paper**/
/*****/
/*Note: adherence = 0 for missed visits*****/
/*%difference and 95%CI based on 500 bootstraps*****/
/*****/

libname cdp "</path/>";

/*A. Unadjusted 1980 analysis*/
%macro partA_unadjusted(outdest = , inset=, titlemain =, nboot = , lib = );
```

```
proc printto print = &outdest;
run;
```

```
%let rawdata = &lib..&inset;
```

```
/******A.1. Unadjusted******/
/*Original coding for missing adherence (missed visits: adherence = 0)*/
```

```
title &titlemain;
```

```
/*Tabulate Unadjusted Probabilities*/
```

```
proc freq data = &rawdata;
    tables old_adhx15bin*dth5 / nopercnt nocol cl binomial riskdiff;
run;
```

```
%mend partA_unadjusted;
```

```
/*B. Adjusted via linear regression, baseline covariates only*/
%macro partA_adjusted(nboot = , outdest = , inset=, titlemain =, lib=, adhvar = );
```

```
proc printto print = &outdest;
run;
```

```
%let rawdata = &lib..&inset;
```

```
/******b. Adjusted: linear regression*****/
title &titlemain;
```

```
data onesample ;
    set &rawdata end = _end_ ;
    retain _id ;
    if _n_ = 1 then _id = 0;
    _id = _id + 1 ;
    if _end_ then do ;
        call symput("nids",trim(left(_id)));
    end;
    label adher= "Adherence";
    adher = -1 ; /* 1st copy: equal to original one */
    output ;
    adher = 0 ; /* 2nd copy: adherence set to 0, outcome to missing */
    &adhvar = 0 ;
    dth5 = . ;
    output ;
    adher = 1 ; /* 3rd copy: adherence set to 1, outcome to missing*/
```

```

        &adhvar = 1 ;
        dth5 = . ;
        output ;
run;

data ids ;
    do bsample = 1 to &nboot ;
        do _id = 1 to &nids ;
            output ;
        end;
    end;
run;

proc surveyselect data= ids
    method = urs
    n= &nids
    seed = 1232
    out = _idsamples (keep = bsample _id numberhits )
    outall noprint ;
strata bsample ;
run;

%do bsample = 0 %to &nboot ;

    %if %eval(&bsample) = 0 %then %do;
        proc reg data = onesample plots=none outest = est0;
            model dth5 = &adhvar
                age_bin nonwhite IRK MI_bin NIHA_bin1 RBW_bin
                CHF ACI AP IC ICIA DIG DIUR AntiArr AntiHyp OralHyp
                CardioM AnyQQS AnySTDep AnyTWave STElev FVEB VCD
                HiHeart HiDiasBP HiSySBP HiBili HiSerChol HiSerTrigly HiSerUric HiSerAlk
                HiPlasUrea HiFastGluc HiOneGluc HiWhiteCell HiNeut HiHemat Cig Inact ;
            output out = predicted_mean0 (keep = id &adhvar adher meanY) p = meanY ;
        run;
        quit;

        data params;
            set est0 ;
            bsample = 0 ;
            keep &adhvar bsample;
        run;

        data predicted_mean;
            set predicted_mean0 ;
            bsample = 0 ;
            numberhits = 1 ;
        run;
    %end;

```



```

proc datasets library = work nolist;
    delete est0 predicted_mean0;
run;
%end;

%else %do;

    data bootsample;
        merge onesample _idsamples (where= (bsample = &bsample ));
        by _id;
    run;

    ods listing select none ;
    proc reg data=bootsample plots = none outest = est1;
        model dth5 = &adhvar
            age_bin nonwhite IRK MI_bin NIHA_bin1 RBW_bin
            CHF ACI AP IC ICIA DIG DIUR AntiArr AntiHyp OralHyp
            CardioM AnyQQS AnySTDep AnyTWave STElev FVEB VCD
            HiHeart HiDiasBP HiSySBP HiBili HiSerChol HiSerTrigly HiSerUric HiSerAlk
            HiPlasUrea HiFastGluc HiOneGluc HiWhiteCell HiNeut HiHemat Cig Inact ;
        freq numberhits;
        output out = predicted_mean1 (keep = bsample id &adhvar adher meanY
numberhits) p = meanY;
    run;
    quit;
    ods listing ;

    data est1;
        set est1;
        bsample = %eval(&bsample);
        keep &adhvar bsample;
    run;

    data params ;
        set params est1;
        by bsample ;
    run;

    data predicted_mean;
        set predicted_mean predicted_mean1;
        by bsample ;
    run;

    proc datasets library = work nolist;
        delete est1 predicted_mean1 bootsample;

```

```

        run;
    %end;

%end;

proc sort data = params;
    by bsample &adhvar;
run;

proc sort data = predicted_mean ;
    by bsample adher;
run;

/*Risk Difference and Confidence Intervals*/
data temp;
    set params (where=(bsample = 0));
    call symput("rd",&adhvar);
run;
proc univariate data=params (where = (bsample > 0)) noprint;
    var &adhvar;
    output out = stderrs
    std = rd_std;
run;
data temp;
    set stderrs;
    call symput("std",rd_std);
run;
data final_1980Bbin;
    mean = &rd;
    std = &std;
    lb = mean - 1.96 * std ;
    ub = mean + 1.96 * std ;
    label lb="95% Lower bound"
    ub="95% Upper bound"
    std="Standard Error"
    ;
run;

/*Standardized risks and confidence intervals*/
proc means data = predicted_mean mean noprint ;
    var meanY;
    by bsample adher;
    freq numberhits ;
    output out = results0 (keep = bsample adher mean ) mean = mean ;
run;

```

```

proc transpose data=results0 out=results1b prefix=Risk_;
    by bsample;
    id adher;
    var mean;
run;

data results3;
    set results1b;
    rd = Risk_1 - Risk_0;
run;

data temp3;
    set results3 (where=(bsample = 0));
    call symput("Ard",rd);
    call symput("ARisk0", Risk_0);
    call symput("ARisk1", Risk_1);
run;
proc univariate data=results3 (where = (bsample > 0)) noprint;
    var Risk_0 Risk_1;
    output out = stderrs3
    std = Risk0_std Risk1_std;
run;

data temp3;
    set stderrs3;
    call symput("ARisk0std",risk0_std);
    call symput("ARisk1std",risk1_std);

run;
data final_predB;
    RDmean = &Ard;
    HiAd = &ARisk0;
    HiAdstd = &ARisk0std;
    HiAdlb = HiAd - 1.96 * HiAdstd ;
    HiAdub = HiAd + 1.96 * HiAdstd ;
    LoAd = &ARisk1;
    LoAdstd = &ARisk1std;
    LoAdlb = LoAd - 1.96 * LoAdstd ;
    LoAdub = LoAd + 1.96 * LoAdstd ;
    label HiAdlb="95% Lower bound >80% Adh"
           HiAdub="95% Upper bound >80% Adh"
           LoAdlb="95% Lower bound <80% Adh"
           LoAdub="95% Upper bound <80% Adh"
    ;
run;

```

```

/*Print Results*/
proc printto print = &outdest;
run;
proc print data = final_1980Bbin label noobs ;
title2 "Bootstrap results using &nboot samples" ;
title3 "Linear Regression, Placebo Arm";
title4 "<80% Adherence vs >=80% Adherence";
      var mean std lb ub;
run;

proc print data = final_predB label noobs ;
title2 "Bootstrap results using &nboot samples" ;
title3 "Linear Regression, Placebo Arm";
title4 "<80% Adherence vs >=80% Adherence";
      var RDmean HiAd HiAdlb HiAdub LoAd LoAdlb LoAdub;
run;

proc datasets library=work nolist;
      delete lds Onesample
      params predicted_mean
      results0 results1b results3
      stderrs temp test _idsamples
      temp2 temp3 stderrs3
      ;
quit;

proc printto;
run;
%let timenow2=%sysfunc(time(), time.);
%let datenow2=%sysfunc(date(), date9.);
%put Part A is Complete;
%put End time is &datenow2 / &timenow2 ;
%put ;

%mend partA_adjusted;

/*****RUN MACROS****/
%let nboot = 50;

%partA_unadjusted( outdest = "PartA.Unadj.final.rtf", inset = binary, titlemain = "Unadjusted, 1980
Replication, Missed Adherence = 0",
      nboot = &nboot, lib = cdp);

/*original 1980 adherence definition*/

```

```
%partA_adjusted( outdest = "PartA.Adj.linear.final.rtf", inset = binary,
                 titlemain = 'Adjusted, 1980 Replication, Missed Adherence = 0', nboot = &nboot, lib=cdp,
                 adhvar = old_adhx15bin);
```

```
/*updated adherence definition*/
```

```
%partA_adjusted( outdest = "PartB.Adj.linear.final.rtf", inset = binary,
                 titlemain = 'Adjusted, 1980 Replication, Missed Adherence = 0', nboot = &nboot, lib=cdp,
                 adhvar = adhx15bin);
```

```

/*****
**Program 3: New Analyses: Baseline Covariates Only**
*****/
/*Note: adherence carried forward 2 missed visits****/
/*%difference and 95%CI based on 500 bootstraps*****/
/*Crude and Standardized estimates*****/
*****/

```

```
libname cdp "</path/>";
```

```
%macro partB_unadjusted(outdest = , inset = , titlemain = , nboot=, lib=);
proc printto print = &outdest;
run;
%let rawdata = &lib..&inset;
```

```

/*****a. Unadjusted*****/
title &titlemain;

```

```
/*Tabulate Unadjusted Probabilities*/
```

```
proc freq data = &rawdata;
    tables adhx15bin*dth5 / nopercnt nocol cl binomial riskdiff;
run;
```

```
%mend partB_unadjusted;
```

```
%macro partB_adjusted(outdest = , inset = , titlemain = , nboot=, lib=, adhvar = );
```

```

/*****b. Adjusted*****/

```

```
%let rawdata = &lib..&inset;
```

```

title &titlemain;

proc format ;
  value adher -1= "Observed"
              0= "Adherence >=80%"
              1= "Adherence <80%"
              ;

run;
data onesample ;
  set &rawdata end = _end_ ;
  label adher= "Adherence";

retain _id ;
if _n_ = 1 then _id = 0;
_id = _id + 1 ;
if _end_ then do ;
  call symput("nids",trim(left(_id)));
end;
adher = -1 ;
  output ;
adher = 0 ;
  &adhvar = 0 ;
  dth5 = . ;
  output ;
adher = 1 ;
  &adhvar = 1 ;
  dth5 = . ;
  output ;
run;
data ids ;
  do bsample = 1 to &nboot;
    do _id = 1 to &nids ;
      output ;
    end;
  end;
run;
proc surveyselect data= ids
  method = urs
  n= &nids
  seed = 1232
  out = _idsamples (keep = bsample _id numberhits )
  outall noprint ;
strata bsample ;
run;

%do bsample = 0 %to &nboot;

  %if %eval(&bsample) = 0 %then %do;

```

```
proc printto print = &outdest;
run;
```

```
/*logistic model for death given adherence*/
proc logistic data = onesample descending;
    model dth5 = &adhvar
    adhpre0bin age_bin nonwhite IRK MI_bin NIHA_bin1 RBW_bin
    CHF ACI AP IC ICIA DIG DIUR AntiArr AntiHyp OralHyp
    CardioM AnyQQS AnySTDep AnyTWave STElev FVEB VCD
    HiHeart HiDiasBP HiSySBP HiBili HiSerChol HiSerTrigly HiSerUric HiSerAlk
    HiPlasUrea HiFastGluc HiOneGluc HiWhiteCell HiNeut HiHemat Cig Inact ;
    output out = predicted_mean0 (keep=adher probY) p = probY ;
run;
data predicted_mean;
    set predicted_mean0 ;
    bsample = 0 ;
    numberhits = 1 ;
run;

proc datasets library = work nolist;
    delete predicted_mean0;
run;
```

```
proc printto ;
run;

    %end;
```

```
%else %do;
    data bootsample;
        merge onesample _idsamples (where = (bsample = &bsample));
        by _id;
    run;

    ods listing select none ;
    proc logistic data=bootsample descending;
        model dth5 = &adhvar
        adhpre0bin age_bin nonwhite IRK MI_bin NIHA_bin1 RBW_bin
        CHF ACI AP IC ICIA DIG DIUR AntiArr AntiHyp OralHyp
        CardioM AnyQQS AnySTDep AnyTWave STElev FVEB VCD
        HiHeart HiDiasBP HiSySBP HiBili HiSerChol HiSerTrigly HiSerUric HiSerAlk
        HiPlasUrea HiFastGluc HiOneGluc HiWhiteCell HiNeut HiHemat Cig Inact;
        output out = predicted_mean1 (keep=bsample numberhits adher probY) p =
probY ;

        freq numberhits ;
        by bsample ;
    run;
    ods listing ;
```

```

        data predicted_mean ;
            set predicted_mean predicted_mean1;
            by bsample ;
        run;

        proc datasets library = work nolist;
            delete predicted_mean1 bootsample;
        run;
    %end;
%end;

proc printto print = &outdest;
run;

proc sort data = predicted_mean ;
    by bsample adher;
run;

/*Risk Difference and Confidence Intervals*/
proc means data=predicted_mean mean noprint;
    var probY ;
    by bsample adher;
    freq numberhits;
    output out = results (keep = bsample adher mean) mean=mean;
run;

proc transpose data=results out = for_diff prefix = Risk_;
    var mean;
    id adher;
    by bsample;
run;
data for_diff;
    set for_diff;
    mean = Risk_1 - Risk_0;
    adher = 2;
    keep bsample adher mean;
run;
proc means data = for_diff (where = (bsample >0)) noprint;
    var mean;
    by adher;
    output out = diffstd (keep = adher std) std = std;
run;
data sample0;
    set for_diff (where = (bsample = 0));
    keep mean;
run;
data final_2014Bbin;

```



```

        merge sample0 diffstd;
        lb = mean - 1.96 * std ;
        ub = mean + 1.96 * std ;
        label lb="95% Lower bound"
              ub="95% Upper bound"
        std="Standard Error"
    ;
run;

/*Print Results*/
proc printto print = &outdest;
run;
proc print data= final_2014Bbin label noobs ;
title2 "95% Confidence Intervals using &nboot samples" ;
title3 "Logistic Regression, Placebo Arm";
title4 "<80% Adherence vs >=80% Adherence";
    var mean std lb ub;
run;
proc datasets library=work nolist;
    delete lds Onesample
    exp_plc plc plc0 plc1
    predicted_mean results
    diffstd for_diff sample0 _idsamples
    ;
quit;

proc printto;
run;

%let timenow2=%sysfunc(time(), time.);
%let datenow2=%sysfunc(date(), date9.);
%put Part B is Complete;
%put End time is &datenow2 / &timenow2 ;
%put ;

%mend partB_adjusted;

/*****RUN MACROS*****/

%let nboot = 50;

%partB_unadjusted( outdest = "PartB.Unadj.final.rtf", inset = binary,
    titlemain = 'Unadjusted, Missed Adherence Carried Forward', nboot = &nboot, lib=cdp);

%partB_adjusted( outdest = "PartB.Adj.logistic.final.rtf", inset = binary,
    titlemain = 'Adjusted, Missed Adherence Carried Forward', nboot = &nboot, lib=cdp,adhvar =
    adhx15bin);

```

```

/*Run logistic regression with original 1980 adherence definition*/
%partB_adjusted( outdest = "PartA.Adj.logistic.final.rtf", inset = binary,
                 titlemain = 'Adjusted, Missed Adherence Carried Forward', nboot = &nboot, lib=cdp,adhvar =
old_adhx15bin);

```

```

/*****/
/**Program 4: New Analyses: Post-randomization Covariates **/
/*****/
/*Note: adherence carried forward 2 missed visits*****/
/**Regimes: Cumulative average >=80% vs *****/
/**cumulative average <80%*****/
/*%difference and 95%CI based on 500 bootstraps*****/
/*IPW for adherence level at time t*****/
/*****/

```

```
libname cdp "</path/>";
```

```
%include "rcspline.sas";
```

```

%macro partC(outdest = , inset = , titlemain = , nboot= , lib=);
title &titlemain;
%let rawdata = &lib..&inset;
%put &rawdata;

```

```

/*Inverse probability weighting for censoring due to change in adherence*/
proc sort data=&rawdata out=onesample;
by id visit;
run;

```

```

data onesample ;
  set onesample end = _end_ ;
  by ID;
  retain _id ;
  if _n_ = 1 then _id = 0;
  if first.id then do;
    _id = _id + 1 ;
  end;
  if _end_ then do ;
    call symput("nids",trim(left(_id)));
  end;

```

```

/*Spline of time*/
%rcspline(visit,0,8,17);
run;

```

```

data ids ;
  do bsample = 1 to &nboot;

```

```

do _id = 1 to &nids ;
  output ;
end;
end;
run;
proc surveyselect data= ids
  method = urs
  n= &nids
  seed = 1232
  out = _idsamples (keep = bsample _id numberhits )
  outall noprint ;
strata bsample ;
run;

%do bsample = 0 %to &nboot;

  %if %eval(&bsample) = 0 %then %do;

proc printto print = &outdest;
run;

/*IPW for adherence at time t - point estimate*/

/*model for adherence measured at time t*/
/*Numerator: Pr(Adh_measured=1 | A_0, Baseline covariates)*/

proc logistic data = onesample (where =( adhbin0 ne . and adhpre0bin ne . and adhr_t1
ne . and visit >0 and
      age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
      NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0 ne . and
      HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
      HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and APO ne . and
      ICO ne . and ICIA0 ne . and DIG0 ne . and DIURO ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
      CardioM0 ne . and AnyQQS0 ne . and AnySTDdep0 ne . and AnyTWave0 ne .
and
      STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and

      NIHAFV_t1 ne . and HiSysBPFV_t1 ne . and HiDiasBPFV_t1 ne . and
HiWhiteCellFV_t1 ne . and HiNeutFV_t1 ne . and HiHematFV_t1 ne . and

```

```

        HiBiliFV_t1 ne . and HiSerCholFV_t1 ne . and HiSerTriglyFV_t1 ne . and
HiSerUricFV_t1 ne . and HiSerAlkFV_t1 ne . and
        HiPlasUreaFV_t1 ne . and HiFastGlucFV_t1 ne . and HiOneGlucFV_t1 ne . and
HiHeartFV_t1 ne . and
        CHFFV_t1 ne . and ACIFV_t1 ne . and APFV_t1 ne . and ICFV_t1 ne . and
ICIAFV_t1 ne . and DIGFV_t1 ne . and DIURFV_t1 ne . and AntiArrFV_t1 ne .
and
        AntiHypFV_t1 ne . and OralHypFV_t1 ne . and CardioMFV_t1 ne . and
AnyQSFV_t1 ne . and AnySTDepFV_t1 ne . and
        AnyTWaveFV_t1 ne . and STElevFV_t1 ne . and FVEBFV_t1 ne . and VCDFV_t1
ne . and
        CIGFV_t1 ne . and INACTFV_t1 ne .)) descending;

model adh_measure = visit visit1 adhbin0 adhr_t1
adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQSQ0 AnySTDep0 AnyTWave0
STElev0 FVEB0 VCD0 CIG0 INACT0 ;
output out = adhmeas_num0 (keep=id visit mesr_0a0) p = mesr_0a0;
run;

/*Denominator: Pr(Adh_measured=1|Baseline covariates, Time-varying covariates)*/

proc logistic data = onesample (where =( adhbin0 ne . and adhpre0bin ne . and adhr_t1
ne . and visit >0 and
        age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
        NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0 ne . and
        HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
        HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and APO ne . and
        ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
        CardioM0 ne . and AnyQSQ0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
        STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and
        NIHAFV_t1 ne . and HiSysBPFV_t1 ne . and HiDiasBPFV_t1 ne . and
HiWhiteCellFV_t1 ne . and HiNeutFV_t1 ne . and HiHematFV_t1 ne . and

```

```

                                HiBiliFV_t1 ne . and HiSerCholFV_t1 ne . and HiSerTriglyFV_t1 ne . and
HiSerUricFV_t1 ne . and HiSerAlkFV_t1 ne . and
                                HiPlasUreaFV_t1 ne . and HiFastGlucFV_t1 ne . and HiOneGlucFV_t1 ne . and
HiHeartFV_t1 ne . and
                                CHFFV_t1 ne . and ACIFV_t1 ne . and APFV_t1 ne . and ICFV_t1 ne . and
                                ICIAFV_t1 ne . and DIGFV_t1 ne . and DIURFV_t1 ne . and AntiArrFV_t1 ne .
and
                                AntiHypFV_t1 ne . and OralHypFV_t1 ne . and CardioMFV_t1 ne . and
AnyQQSFV_t1 ne . and AnySTDepFV_t1 ne . and
                                AnyTWaveFV_t1 ne . and STElevFV_t1 ne . and FVEBFV_t1 ne . and VCDFV_t1
ne . and
                                CIGFV_t1 ne . and INACTFV_t1 ne .)) descending ;

                                model adh_measure = visit visit1 adhbin0 adhr_t1
                                adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

                                NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
                                HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
                                HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
                                ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
                                CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
                                STElev0 FVEB0 VCD0 CIG0 INACT0

                                NIHAfV_t1 HiSysBPFV_t1 HiDiasBPFV_t1 HiWhiteCellfV_t1 HiNeutfV_t1
HiHematfV_t1
                                HiBiliFV_t1 HiSerCholFV_t1 HiSerTriglyFV_t1 HiSerUricFV_t1 HiSerAlkFV_t1
                                HiPlasUreaFV_t1 HiFastGlucFV_t1 HiOneGlucFV_t1 HiHeartFV_t1
                                CHFFV_t1 ACIFV_t1 APFV_t1 ICFV_t1 ICIAFV_t1 DIGFV_t1 DIURFV_t1
AntiArrFV_t1
                                AntiHypFV_t1 OralHypFV_t1 CardioMFV_t1 AnyQQSFV_t1 AnySTDepFV_t1
                                AnyTWaveFV_t1 STElevFV_t1 FVEBFV_t1 VCDFV_t1
                                CIGFV_t1 INACTFV_t1 ;
                                output out = adhmeas_dnom0 (keep=id visit mesr_wa0) p = mesr_wa0;
run;

/*model for adherence at time t, given adherence measured*/

/*Numerator: Pr(A_t=1|A_0, Baseline covariates)*/

proc logistic data=onesample (where =( adhbin0 ne . and adhr_t ne . and adhr_t1 ne .
and adhpre0bin ne . and visit >0 and
                                age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
                                NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0
                                ne . and
                                HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and

```

```

                                HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and AP0 ne . and
                                ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
                                CardioM0 ne . and AnyQSQ0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
                                STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and

                                NIHAFV ne . and HiSysBPFV ne . and HiDiasBPFV ne . and HiWhiteCellFV ne .
and HiNeutFV ne . and HiHematFV ne . and
                                HiBiliFV ne . and HiSerCholFV ne . and HiSerTriglyFV ne . and HiSerUricFV ne .
and HiSerAlkFV ne . and
                                HiPlasUreaFV ne . and HiFastGlucFV ne . and HiOneGlucFV ne . and
HiHeartFV ne . and
                                CHFFV ne . and ACIFV ne . and APFV ne . and ICFV ne . and ICIAFV ne . and
DIGFV ne . and DIURFV ne . and AntiArrFV ne . and
                                AntiHypFV ne . and OralHypFV ne . and CardioMFV ne . and AnyQSQFV ne .
and AnySTDepFV ne . and
                                AnyTWaveFV ne . and STElevFV ne . and FVEBFV ne . and VCDFV ne . and
CIGFV ne . and INACTFV ne .)) descending;

                                model adhr_t = visit visit1 adhbin0 adhr_t1
                                adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

                                NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
                                HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
                                HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO AP0
                                ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
                                CardioM0 AnyQSQ0 AnySTDep0 AnyTWave0
                                STElev0 FVEB0 VCD0 CIG0 INACT0 ;
                                output out = censadh_num0 (keep=id visit punc_0a0) p = punc_0a0;
run;

/*Denominator: Pr(A_t=1|A_0, Baseline covariates, Time-varying covariates)*/

proc logistic data=onesample (where =( adhbin0 ne . and adhr_t ne . and adhr_t1 ne .
and adhpre0bin ne . and visit >0 and
                                age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
                                NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0
                                ne . and
                                HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
                                HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and AP0 ne . and
                                ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and

```

```

and CardioM0 ne . and AnyQQS0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and NIHAFV ne . and HiSysBPFV ne . and HiDiasBPFV ne . and HiWhiteCellFV ne .
and HiNeutFV ne . and HiHematFV ne . and
HiBiliFV ne . and HiSerCholFV ne . and HiSerTriglyFV ne . and HiSerUricFV ne .
and HiSerAlkFV ne . and
HiPlasUreaFV ne . and HiFastGlucFV ne . and HiOneGlucFV ne . and
HiHeartFV ne . and
CHFFV ne . and ACIFV ne . and APFV ne . and ICFV ne . and ICIAFV ne . and
DIGFV ne . and DIURFV ne . and AntiArrFV ne . and
AntiHypFV ne . and OralHypFV ne . and CardioMFV ne . and AnyQQSFV ne .
and AnySTDepFV ne . and
AnyTWaveFV ne . and STElevFV ne . and FVEBFV ne . and VCDFV ne . and
CIGFV ne . and INACTFV ne .)) descending;

```

```

model adhr_t = visit visit1 adhbin0 adhr_t1
adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

```

```

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STElev0 FVEB0 VCD0 CIG0 INACT0

```

```

NIHAFV HiSysBPFV HiDiasBPFV HiWhiteCellFV HiNeutFV HiHematFV
HiBiliFV HiSerCholFV HiSerTriglyFV HiSerUricFV HiSerAlkFV
HiPlasUreaFV HiFastGlucFV HiOneGlucFV HiHeartFV
CHFFV ACIFV APFV ICFV ICIAFV DIGFV DIURFV AntiArrFV
AntiHypFV OralHypFV CardioMFV AnyQQSFV AnySTDepFV
AnyTWaveFV STElevFV FVEBFV VCDFV
CIGFV INACTFV;
output out = censadh_dnom0 (keep = id visit punc_wa0) p = punc_wa0;

```

```
run;
```

```

proc sort data=adhmeas_num0;
  by ID visit;
proc sort data=adhmeas_dnom0;
  by ID visit;
proc sort data=censadh_num0;
  by ID visit;
proc sort data=censadh_dnom0;
  by ID visit;

```

```

proc sort data=onesample;
    by ID visit;
data main_w6;
    merge onesample censadh_num0 censadh_dnom0 adhmeas_num0
adhmeas_dnom0 ;
    by ID visit;

/* variables ending with _0 refer to the numerator of the weights
Variables ending with _w refer to the denominator of the weights */

if first.id then do;
    k1_0=1;
    k1_w=1;
    m1_0=1;
    m1_w=1;
end;
retain k1_0 k1_w m1_0 m1_w;

if adhr_t ne . then do;

    if mesr_0a0 = . then mesr_0a0 =1;
    if mesr_wa0 = . then mesr_wa0 =1;
    m1_0=m1_0*mesr_0a0;
    m1_w=m1_w*mesr_wa0;

    if adhr_t = 0 then do;
        if punc_0a0 = . then punc_0a0 =0;
        if punc_wa0 = . then punc_wa0 =0;

        k1_0=k1_0*(1-punc_0a0);
        k1_w=k1_w*(1-punc_wa0);
    end;
    else if adhr_t = 1 then do;
        if punc_0a0 = . then punc_0a0 =1;
        if punc_wa0 = . then punc_wa0 =1;
        k1_0=k1_0*(punc_0a0);
        k1_w=k1_w*(punc_wa0);
    end;

end;

else if adhr_t = . then do;
    if mesr_0a0 = . then mesr_0a0 =0;
    if mesr_wa0 = . then mesr_wa0 =0;
    m1_0=m1_0*(1-mesr_0a0);
    m1_w=m1_w*(1-mesr_wa0);

    k1_0 = k1_0*1;
    k1_w = k1_w*1;

```



```

        end;

        stabw_a=(k1_0)/(k1_w);
        nstabw_a=1/(k1_w);

        stabw_m=(m1_0)/(m1_w);
        nstabw_m=1/(m1_w);

        stabw =stabw_a*stabw_m;
        nstabw =nstabw_a*nstabw_m;

run;
proc printto print = &outdest;
run;
proc means data=main_w6 n mean std min max p95 p99 nmiss;
        var nstabw stabw nstabw_a stabw_a nstabw_m stabw_m;
        title 'weights, all';
run;
proc printto print = &outdest;
run;
proc freq data=main_w6 nlevels;
        where stabw ne .;
        tables ID /noprint;
        title 'weights, all';
run;

data endfup;
        set main_w6 end = _end_ ;
        by ID;
        retain _id ;
        if _n_ = 1 then _id = 0;
        if last.id then do;
                _id = _id + 1 ;
                output;
        end;
run;

proc printto print = &outdest;
run;
proc means data=endfup n mean std min max p95 p99 p1 nmiss;
        var nstabw stabw nstabw_a stabw_a nstabw_m stabw_m;
        title 'weights, end of follow-up';
run;

proc freq data=endfup nlevels;
        where stabw ne .;
        tables ID /noprint;
        title 'weights, end of follow-up';

```

```

run;

/*for truncation*/
proc means data=endfup n mean std min max p95 p99 nmiss noprint;
    var stabw;
    title 'stabilized weights, end of follow-up';
    output out=pctl (keep = p99) p99 = p99 ;
run;
proc means data=endfup p99 noprint;
    var nstabw;
    title 'stabilized weights, end of follow-up';
    output out=pctl_n (keep = p99) p99 = p99 ;
run;

data temp;
    set pctl;
    call symput ('cutoff', p99);
run;
data temp_n;
    set pctl_n;
    call symput ('cutoff_n', p99);
run;

data trunc;
set endfup;
    stabw1 = stabw;
    if stabw > %sysevalf(&cutoff) then do;
        stabw1 = %sysevalf(&cutoff);
    end;
    nstabw1 = nstabw;
    if nstabw > %sysevalf(&cutoff_n) then do;
        nstabw1 = %sysevalf(&cutoff_n);
    end;
run;
proc means data=trunc n mean std min max p95 p99 nmiss;
    var stabw stabw1 nstabw nstabw1;
    title 'stabilized weights, end of follow-up';
run;

/*Weighted regression model*/
/*Pr(Y=1 | Adherence at last visit, Baseline covariates)*/
proc logistic data = trunc descending ;
ods output ParameterEstimates = PE;
    model dth5 = adhx15bin
    adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

```

```

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STelev0 FVEB0 VCD0 CIG0 INACT0 ;
weight stabw1;
title 'Main Weighted model';
run;

proc sql noprint;
select ESTIMATE FORMAT =16.12 INTO: IBC_ESTIMATE separated by ' ' from pe;
quit;
proc sql noprint;
select variable INTO: model separated by ' ' from PE;
quit;

proc means sum noprint data = pe;
var df;
output out = nobs (drop = _type_ _freq_ where=(_stat_ ="N"));
run;
proc sql noprint;
select df into:nvar separated by ' ' from nobs;
quit;

data rq1 (keep = p0 p1 numberhits);
set trunc;
array var{&nvar} &model;
array coef{&nvar} (&ibc_estimate);

intercept = 1;
numberhits = 1;

xbeta0 = 0;
xbeta1 = 0;
do i = 1 to dim(var);
    adhx15bin =0;
    xbeta0 = xbeta0 + coef[i] *var[i];

    adhx15bin = 1;
    xbeta1 = xbeta1 + coef[i]*var[i];
end;

p0 = 1/(1+exp(-xbeta0));
p1 = 1/(1+exp(-xbeta1));

output;
run;

```

```

proc means data = rq1 mean noprint;
    var p0 p1;
    freq numberhits;
    output out = mean_1 (drop = _type_ _freq_) mean(p0 p1) = ;
run;

data mean_1;
    set mean_1;
    label p0 = "Cumulative Adherence < 80%"
           p1 = "Cumulative Adherence >= 80%";
    bsample = &bsample;
run;

data means_all;
    set mean_1;
run;

proc datasets library = work nolist;
    delete censadh_num0 censadh_dnom0 temp pctl rq1 pe mean_1;
run;
%symdel cutoff;

```

```

proc printto ;
run;

```

```

%end;

```

```

%else %do;

```

```

    title "Bootstrap number &bsample";

```

```

proc sort data = onesample ;

```

```

    by _id;

```

```

run;

```

```

data bootsample;

```

```

    merge onesample _idsamples (where = (bsample = &bsample));

```

```

    by _id;

```

```

run;

```

```

proc sort data = bootsample sortsize=5G ;

```

```

    by id visit ;

```

```

run;

```

```

/*IPW for adherence at time t - bootstrap sample*/

```

```

/*model for adherence measured at time t*/
/*Numerator: Pr(Adh_measured=1|A_0, Baseline covariates)*/

proc logistic data = bootsample (where =( adhbin0 ne . and adhpre0bin ne . and adhr_t1
ne . and visit >0 and
                                age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
                                NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0      ne . and
                                HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
                                HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and AP0 ne . and
                                ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
                                CardioM0 ne . and AnyQQS0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
                                STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and

                                NIHAFV_t1 ne . and HiSysBPFV_t1 ne . and HiDiasBPFV_t1 ne . and
HiWhiteCellFV_t1 ne . and HiNeutFV_t1 ne . and HiHematFV_t1 ne . and
                                HiBiliFV_t1 ne . and HiSerCholFV_t1 ne . and HiSerTriglyFV_t1 ne . and
HiSerUricFV_t1 ne . and HiSerAlkFV_t1 ne . and
                                HiPlasUreaFV_t1 ne . and HiFastGlucFV_t1 ne . and HiOneGlucFV_t1 ne . and
HiHeartFV_t1 ne . and
                                CHFFV_t1 ne . and ACIFV_t1 ne . and APFV_t1 ne . and ICFV_t1 ne . and
ICIAFV_t1 ne . and DIGFV_t1 ne . and DIURFV_t1 ne . and AntiArrFV_t1 ne .
and
                                AntiHypFV_t1 ne . and OralHypFV_t1 ne . and CardioMFV_t1 ne . and
AnyQQSFV_t1 ne . and AnySTDepFV_t1 ne . and
                                AnyTWaveFV_t1 ne . and STElevFV_t1 ne . and FVEBFV_t1 ne . and VCDFV_t1
ne . and
                                CIGFV_t1 ne . and INACTFV_t1 ne .)) descending;

model adh_measure = visit visit1 adhbin0 adhr_t1
adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO AP0
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STElev0 FVEB0 VCD0 CIG0 INACT0 ;
freq numberhits ;
output out = adhmeasbootnum0 (keep=id visit mesr_0a0) p = mesr_0a0;

run;

```

```
/*Denominator: Pr(Adh_measured=1|Baseline covariates, Time-varying covariates)*/

proc logistic data = bootsample (where =( adhbin0 ne . and adhpre0bin ne . and
adhr_t1 ne . and visit >0 and
                                age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
                                NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0      ne . and
                                HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
                                HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and APO ne . and
                                ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
                                CardioM0 ne . and AnyQQS0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
                                STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and

                                NIHAFV_t1 ne . and HiSysBPFV_t1 ne . and HiDiasBPFV_t1 ne . and
HiWhiteCellFV_t1 ne . and HiNeutFV_t1 ne . and HiHematFV_t1 ne . and
                                HiBiliFV_t1 ne . and HiSerCholFV_t1 ne . and HiSerTriglyFV_t1 ne . and
HiSerUricFV_t1 ne . and HiSerAlkFV_t1 ne . and
                                HiPlasUreaFV_t1 ne . and HiFastGlucFV_t1 ne . and HiOneGlucFV_t1 ne . and
HiHeartFV_t1 ne . and
                                CHFFV_t1 ne . and ACIFV_t1 ne . and APFV_t1 ne . and ICFV_t1 ne . and
ICIAFV_t1 ne . and DIGFV_t1 ne . and DIURFV_t1 ne . and AntiArrFV_t1 ne .
and
                                AntiHypFV_t1 ne . and OralHypFV_t1 ne . and CardioMFV_t1 ne . and
AnyQQSFV_t1 ne . and AnySTDepFV_t1 ne . and
                                AnyTWaveFV_t1 ne . and STElevFV_t1 ne . and FVEBFV_t1 ne . and VCDFV_t1
ne . and
                                CIGFV_t1 ne . and INACTFV_t1 ne .)) descending ;

model adh_measure = visit visit1 adhbin0 adhr_t1
adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STElev0 FVEB0 VCD0 CIG0 INACT0
```

```

NIHAFV_t1 HiSysBPFV_t1 HiDiasBPFV_t1 HiWhiteCellFV_t1 HiNeutFV_t1
HiHematFV_t1
HiBiliFV_t1 HiSerCholFV_t1 HiSerTriglyFV_t1 HiSerUricFV_t1 HiSerAlkFV_t1
HiPlasUreaFV_t1 HiFastGlucFV_t1 HiOneGlucFV_t1 HiHeartFV_t1
CHFFV_t1 ACIFV_t1 APFV_t1 ICFV_t1 ICIAFV_t1 DIGFV_t1 DIURFV_t1
AntiArrFV_t1
AntiHypFV_t1 OralHypFV_t1 CardioMFV_t1 AnyQQSFV_t1 AnySTDepFV_t1
AnyTWaveFV_t1 STElevFV_t1 FVEBFV_t1 VCDFV_t1
CIGFV_t1 INACTFV_t1 ;
freq numberhits ;
output out = adhmeasbootden0 (keep=id visit mesr_wa0) p = mesr_wa0;
run;

/*model for adherence at time t, given adherence measured*/

/*Numerator: Pr(A_t=1|A_0, Baseline covariates)*/

proc logistic data=bootsample (where =( adhbin0 ne . and adhr_t ne . and adhr_t1 ne .
and adhpre0bin ne . and visit >0 and
age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0 ne . and
HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and AP0 ne . and
ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
CardioM0 ne . and AnyQQS0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
STElev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and
NIHAFV ne . and HiSysBPFV ne . and HiDiasBPFV ne . and HiWhiteCellFV ne .
and HiNeutFV ne . and HiHematFV ne . and
HiBiliFV ne . and HiSerCholFV ne . and HiSerTriglyFV ne . and HiSerUricFV ne .
and HiSerAlkFV ne . and
HiPlasUreaFV ne . and HiFastGlucFV ne . and HiOneGlucFV ne . and
HiHeartFV ne . and
CHFFV ne . and ACIFV ne . and APFV ne . and ICFV ne . and ICIAFV ne . and
DIGFV ne . and DIURFV ne . and AntiArrFV ne . and
AntiHypFV ne . and OralHypFV ne . and CardioMFV ne . and AnyQQSFV ne .
and AnySTDepFV ne . and
AnyTWaveFV ne . and STElevFV ne . and FVEBFV ne . and VCDFV ne . and
CIGFV ne . and INACTFV ne .)) descending;

model adhr_t = visit visit1 adhbin0 adhr_t1

```

```

adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STelev0 FVEB0 VCD0 CIG0 INACT0 ;
freq numberhits ;
output out = cnsadhbootnum0 (keep=id visit punc_0a0) p = punc_0a0;
run;

/*Denominator: Pr(A_t=1|A_0, Baseline covariates, Time-varying covariates)*/

proc logistic data=bootsample (where =( adhbin0 ne . and adhr_t ne . and adhr_t1 ne .
and adhpre0bin ne . and visit >0 and
age_bin ne . and nonwhite ne . and IRK ne . and MI_bin ne . and RBW_bin ne .
and
NIHA_bin0 ne . and HiSysBP0 ne . and HiDiasBP0 ne . and HiWhiteCell0 ne .
and HiNeut0 ne . and HiHemat0 ne . and
HiBili0 ne . and HiSerChol0 ne . and HiSerTrigly0 ne . and HiSerUric0 ne . and
HiSerAlk0 ne . and HiPlasUrea0 ne . and
HiFastGluc0 ne . and HiOneGluc0 ne . and HiHeart0 ne . and CHF0 ne . and
ACIO ne . and APO ne . and
ICO ne . and ICIA0 ne . and DIG0 ne . and DIUR0 ne . and AntiArr0 ne . and
AntiHyp0 ne . and OralHyp0 ne . and
CardioM0 ne . and AnyQQS0 ne . and AnySTDep0 ne . and AnyTWave0 ne .
and
STelev0 ne . and FVEB0 ne . and VCD0 ne . and CIG0 ne . and INACT0 ne .
and
NIHAFV ne . and HiSysBPFV ne . and HiDiasBPFV ne . and HiWhiteCellFV ne .
and HiNeutFV ne . and HiHematFV ne . and
HiBiliFV ne . and HiSerCholFV ne . and HiSerTriglyFV ne . and HiSerUricFV ne .
and HiSerAlkFV ne . and
HiPlasUreaFV ne . and HiFastGlucFV ne . and HiOneGlucFV ne . and
HiHeartFV ne . and
CHFFV ne . and ACIFV ne . and APFV ne . and ICFV ne . and ICIAFV ne . and
DIGFV ne . and DIURFV ne . and AntiArrFV ne . and
AntiHypFV ne . and OralHypFV ne . and CardioMFV ne . and AnyQQSFV ne .
and AnySTDepFV ne . and
AnyTWaveFV ne . and STElevFV ne . and FVEBFV ne . and VCDFV ne . and
CIGFV ne . and INACTFV ne .)) descending;

model adhr_t = visit visit1 adhbin0 adhr_t1
adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin

```



```

NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0
HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
IC0 ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
STelev0 FVEB0 VCD0 CIG0 INACT0

```

```

NIHAFV HiSysBPFV HiDiasBPFV HiWhiteCellFV HiNeutFV HiHematFV
HiBiliFV HiSerCholFV HiSerTriglyFV HiSerUricFV HiSerAlkFV
HiPlasUreaFV HiFastGlucFV HiOneGlucFV HiHeartFV
CHFFV ACIFV APFV ICFV ICIAFV DIGFV DIURFV AntiArrFV
AntiHypFV OralHypFV CardioMFV AnyQQSFV AnySTDepFV
AnyTWaveFV STElevFV FVEBFV VCDFV
CIGFV INACTFV;

```

```

freq numberhits ;

```

```

output out = cnsadhbootden0 (keep = id visit punc_wa0) p = punc_wa0;

```

```

run;

```

```

proc sort data=adhmeasbootnum0;

```

```

by ID visit;

```

```

proc sort data=adhmeasbootden0;

```

```

by ID visit;

```

```

proc sort data=cnsadhbootnum0;

```

```

by ID visit;

```

```

proc sort data=cnsadhbootden0;

```

```

by ID visit;

```

```

proc sort data = bootsample sortsize=5G ;

```

```

by id visit ;

```

```

run;

```

```

data weighted;

```

```

merge bootsample cnsadhbootnum0 cnsadhbootden0 adhmeasbootnum0

```

```

adhmeasbootden0;

```

```

by ID visit;

```

```

/* variables ending with _0 refer to the numerator of the weights

```

```

Variables ending with _w refer to the denominator of the weights */

```

```

if first.id then do;

```

```

k1_0=1;

```

```

k1_w=1;

```

```

m1_0=1;

```

```

m1_w=1;

```

```

end;

```

```

retain k1_0 k1_w m1_0 m1_w;

```

```

if adhr_t ne . then do;

    if mesr_0a0 = . then mesr_0a0 =1;
    if mesr_wa0 = . then mesr_wa0 =1;
    m1_0=m1_0*mesr_0a0;
    m1_w=m1_w*mesr_wa0;

    if adhr_t = 0 then do;
        if punc_0a0 = . then punc_0a0 =0;
        if punc_wa0 = . then punc_wa0 =0;

        k1_0=k1_0*(1-punc_0a0);
        k1_w=k1_w*(1-punc_wa0);
    end;
    else if adhr_t = 1 then do;
        if punc_0a0 = . then punc_0a0 =1;
        if punc_wa0 = . then punc_wa0 =1;
        k1_0=k1_0*(punc_0a0);
        k1_w=k1_w*(punc_wa0);
    end;

end;

else if adhr_t = . then do;
    if mesr_0a0 = . then mesr_0a0 =0;
    if mesr_wa0 = . then mesr_wa0 =0;
    m1_0=m1_0*(1-mesr_0a0);
    m1_w=m1_w*(1-mesr_wa0);

    k1_0 = k1_0*1;
    k1_w = k1_w*1;
end;

stabw_a=(k1_0)/(k1_w);
nstabw_a=1/(k1_w);

stabw_m=(m1_0)/(m1_w);
nstabw_m=1/(m1_w);

stabw =stabw_a*stabw_m;
nstabw =nstabw_a*nstabw_m;

run;

data endfup_boot;
    set weighted end = _end_ ;
    by ID;
    retain _id ;
    if _n_ = 1 then _id = 0;

```

```

        if last.id then do;
            _id = _id + 1 ;
            output;
        end;
run;

proc means data=endfup_boot p99 noprint;
    var stabw;
    title 'stabilized weights, end of follow-up';
    output out=pctl_boot (keep = p99) p99 = p99 ;
run;
proc means data=endfup_boot p99 noprint;
    var nstabw;
    title 'unstabilized weights, end of follow-up';
    output out=pctl_boot_n (keep = p99) p99 = p99 ;
run;

data temp_boot;
    set pctl_boot;
    call symput ('cutoff', p99);
run;

data temp_boot_n;
    set pctl_boot_n;
    call symput ('cutoff_n', p99);
run;

data trunc_boot;
    set endfup_boot;
    stabw1 = stabw;
    if stabw > %sysevalf(&cutoff) then do;
        stabw1 = %sysevalf(&cutoff);
    end;
    nstabw1 = nstabw;
    if nstabw > %sysevalf(&cutoff_n) then do;
        nstabw1 = %sysevalf(&cutoff_n);
    end;
run;

/*Weighted regression model*/
/*Pr(Y=1| Cumulative Adherence to last visit, Baseline covariates)*/
proc logistic data = trunc_boot descending;
ods output ParameterEstimates = pe_boot;
    model dth5 = adhx15bin
        adhpre0bin age_bin nonwhite IRK MI_bin RBW_bin
        NIHA_bin0 HiSysBP0 HiDiasBP0 HiWhiteCell0 HiNeut0 HiHemat0
        HiBili0 HiSerChol0 HiSerTrigly0 HiSerUric0 HiSerAlk0 HiPlasUrea0

```

```

                HiFastGluc0 HiOneGluc0 HiHeart0 CHF0 ACIO APO
                ICO ICIA0 DIG0 DIUR0 AntiArr0 AntiHyp0 OralHyp0
                CardioM0 AnyQQS0 AnySTDep0 AnyTWave0
                STElev0 FVEB0 VCD0 CIG0 INACT0 ;
weight stabw1;
freq numberhits;
run;

proc sql noprint;
    select ESTIMATE FORMAT =16.12 INTO: IBC_ESTIMATE separated by ' ' from
pe_boot;
quit;
proc sql noprint;
    select variable INTO: model separated by ' ' from pe_boot;
quit;

proc means sum noprint data = pe_boot;
    var df;
    output out = nobs_boot (drop = _type_ _freq_ where=(_stat_ ="N"));
run;
proc sql noprint;
    select df into:nvar separated by ' ' from nobs_boot;
quit;

data rq1_boot (keep = p0 p1 numberhits);
    set trunc_boot;
    array var{&nvar} &model;
    array coef{&nvar} (&ibc_estimate);

    intercept = 1;

    xbeta0 = 0;
    xbeta1 = 0;
    do i = 1 to dim(var);
        adhx15bin =0;
        xbeta0 = xbeta0 + coef[i] *var[i];

        adhx15bin = 1;
        xbeta1 = xbeta1 + coef[i]*var[i];
    end;

    p0 = 1/(1+exp(-xbeta0));
    p1 = 1/(1+exp(-xbeta1));

    output;
run;

proc means data = rq1_boot mean noprint;

```

```

        var p0 p1;
        freq numberhits;
        output out = mean_boot (drop = _type_ _freq_) mean(p0 p1) = ;
run;

data mean_boot;
    set mean_boot;
    label p0 = "Cumulative Adherence < 80%"
          p1 = "Cumulative Adherence >= 80%";
    bsample = &bsample;
run;

data means_all;
    set means_all mean_boot;
    by bsample;
run;

proc datasets library = work nolist;
    delete mean_boot bootsample weighted cnsadhbootnum0 cnsadhbootden0
endfup_boot trunc_boot temp_boot pctl_boot pe_boot rq1_boot;
run;
    %symdel cutoff;
%end;
%end;

proc printto print = &outdest;
run;

title "Summary";

proc sort data=means_all;
by bsample;
run;

/*Risk Difference and Confidence Intervals*/
data rds ;
    set means_all;
    risk_diff = p1-p0;
    keep bsample risk_diff;
run;

proc means data = rds (where = (bsample > 0)) noprint ;
    var risk_diff;
    output out=diffstd (keep = std) std = std ;
run;

data sample0;
    set rds (where=(bsample = 0));

```

```

        keep risk_diff;
run;

data final_2014C;
    merge sample0 diffstd;
    lb = risk_diff - 1.96 * std ;
    ub = risk_diff + 1.96 * std ;
    label lb="95% Lower bound"
           ub="95% Upper bound"
           std="Standard Error"
    ;
run;

proc printto print = &outdest;
run;
proc print data= final_2014C label noobs ;
title &titlemain;
title2 'Adjusted Risk Difference, Logistic regression';
title3 "95% Confidence Intervals using &nboot samples" ;
title4 "<80% Adherence vs >=80% Adherence";
    var risk_diff std lb ub;
run;

proc datasets library=work nolist;
    delete Diffstd Ids Main_w6 Onesample
           results sample0 weighted _idsamples means_all rds
    ;
quit;

proc printto;
run;
%let timenow2=%sysfunc(time(), time.);
%let datenow2=%sysfunc(date(), date9.);
%put Part C is complete;
%put End time is &datenow2 / &timenow2 ;
%put Program is complete;
%put ;

%mend partC;

/**RUN MACROS**/

%let nboot = 50;
%partC( outdest = "PartC.cumA.final.rtf", inset = censwt_ag,
        titlemain = 'Adherence at time t: IPW Adjusted, Missed Adherence Carried Forward', nboot =
        &nboot, lib=cdp);

```

Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials

Manuscript ID:	CT-15-0120.R3	
Funding Information:	U.S. Department of Health and Human Services > National Institutes of Health R01 AI102634 R01 HL080644	
Submitting Author:	<div> <input type="text" value="Murray, Eleanor J"/> <input checked="" type="checkbox"/> Save <input type="text" value="Murray, Eleanor J (proxy)"/> </div> <hr/> <ul style="list-style-type: none"> <i>primary affiliation</i> Harvard TH Chan School of Public Health - Epidemiology 677 Huntington Avenue Boston Massachusetts 02115 United States T: 617-432-1539 	
Authors & Institutions:	<div> <input type="text" value="Murray, Eleanor J proxy"/> <ul style="list-style-type: none"> Harvard TH Chan School of Public Health - Epidemiology 677 Huntington Avenue , Boston, Massachusetts 02115 United States Harvard TH Chan School of Public Health - Biostatistics Boston, Massachusetts United States </div> <div> <input type="text" value="Hernan, Miguel proxy"/> <ul style="list-style-type: none"> Harvard School of Public Health - Epidemiology Department of Epidemiology 677 Huntington Avenue, Boston, Massachusetts 02115 United States Harvard TH Chan School of Public Health - Biostatistics Boston, Massachusetts United States Harvard-MIT Division of Health Sciences and Technology Boston, Massachusetts United States </div>	
Contact Author (populates the ##PROLE_AUTHOR_...## e-mail tags):	<div> <input type="text" value="Murray, Eleanor J"/> <input checked="" type="checkbox"/> Save </div>	Current Contact Author: Murray, Eleanor J (proxy)
Running Head:	Call for better per-protocol estimates	
Keywords:	per-protocol effect ✖, intention-to-treat effect ✖, inverse probability weighting ✖, coronary drug project ✖, adherence ✖	